

**Möglichkeiten der Messung von Einstellungen
als theoretisch relevanter Variablen für die
Erforschung der Entstehung und Wirkung
von Aussagen der Massenmedien**

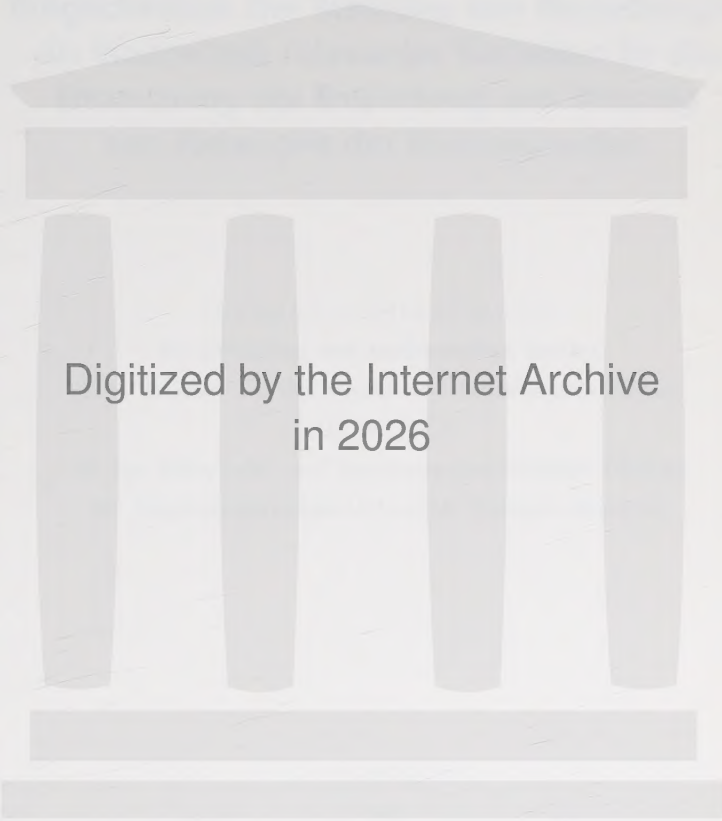
von

Diplom-Volkswirt Manfred Stosberg

aus

Hückeswagen/Rhld.

1969



Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41546218_0004

Möglichkeiten der Messung von Einstellungen als theoretisch relevanter Variablen für die Erforschung der Entstehung und Wirkung von Aussagen der Massenmedien

Inaugural-Dissertation
zur Erlangung des akademischen Grades
eines Doktors der Wirtschafts- und Sozialwissenschaften
(Dr. rer. pol.)
an der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Friedrich-Alexander-Universität Erlangen-Nürnberg

vorgelegt
von
Diplom-Volkswirt Manfred Stosberg
aus
Hückeswagen/Rhld.



Referent: Professor Dr. K. G. Specht

Korreferent: Professor Dr. F. Ronneberger

Tag der mündlichen Prüfung: 23. Juli 1969

INHALTSVERZEICHNIS

	Seite
Möglichkeiten der Messung von Einstellungen als theoretisch relevanter Variablen für die Erforschung der Entstehung und Wirkung von Aussagen der Massenmedien	
Vorwort	1
1 Einleitung	2
2 Theoriebildung und Messung	7
21 Das Ziel der empirischen Wissenschaften	7
22 Die formale Struktur voll entwickelter Theorien	11
23 Zum Stand der Theoriebildung und -prüfung in den Sozialwissenschaften	16
24 Die Bedeutung der Einstellungsmessung für die Bildung und Prüfung von Hypothesen zur Erklärung von Massenkommunikationsprozessen	24
241 Die wichtigsten Vorteile des Ein- stellungsbegriffes	24
242 Einstellungen als Determinanten und Re- sultanten eines theoretischen Konzeptes zur Erklärung von Massenkommunikations- prozessen	34
3 Grundlagen der Messung	39
31 Definitionsversuche: Messung und Skalierung	39
32 Das Niveau der Messung	44
33 Arten der Messung	57
331 Messung per definitionem	58
332 Grundlegende Messung (fundamental measurement)	60
3321 Die Bestimmung der Reliabilität	63
3322 Die Bestimmung der Validität	68
333 Abgeleitete Messung (derived measurement)	77

	Seite
4 Techniken der Skalenkonstruktion	80
41 Vorbemerkungen	80
42 Einteilung der Skalierungstechniken	84
421 Personen-orientiertes Messen	85
422 Indikator-orientiertes Messen	87
423 Reaktions-orientiertes Messen	92
43 Erläuterung und Kritik der Techniken	95
431 Die "Judgment"-Techniken	95
4311 Die Technik des Paarvergleichs	95
4312 Die Technik der gleicherscheinenden Abstände	121
4313 Die Technik der nachträglich be- stimmten Abstände	139
432 Die Technik der summierten Einschätzungen	153
433 Die Technik der Skalogramm-Analyse	168
434 Die Diskriminationstechnik	190
435 Die Technik der transferierten Einschätzungen	200
436 Die UPRO-Technik	221
437 Die Technik der Analyse latenter Strukturen	225
5 Einige Bemerkungen zum gegenwärtigen Stand der Konstruktion von Einstellungsskalen	250

LITERATURVERZEICHNIS

V o r w o r t

Mein Dank soll an dieser Stelle denen gelten, die am Zustandekommen der hier vorgelegten Arbeit Anteil hatten. Diese Dissertation wurde im Rahmen eines größeren Forschungsprojektes zu Problemen der Massenkommunikation unter der Leitung von Herrn Professor Dr. K.G. Specht angefertigt. Herrn Professor Specht habe ich besonders für die verständnisvolle Betreuung und Unterstützung meiner Arbeit zu danken. Ebenso gilt mein Dank Herrn Professor Dr. F. Rönneberger für wertvolle Hinweise und kritische Stellungnahmen.

Meinen Kollegen Herrn Dr. H. Bessler und Herrn Dr. F. Eledjian danke ich für die Anregungen, die sich aus der gemeinsamen Arbeit an dem genannten Forschungsprojekt ergeben haben. Nicht zuletzt möchte ich auch meiner Frau danken, die mir - auch als Kollegin - bei vielen Schwierigkeiten geholfen hat. Auf die Arbeiten meiner Kollegen wird im folgenden wiederholt Bezug genommen.¹⁾

M. S.

1) Vgl. BESSLER, H.: Zur Theoriebildung in der Massenkommunikationsforschung. Ein theoretischer Ansatz und Möglichkeiten der Lösung eines damit verbundenen Meßproblems mit Techniken der Aussagenanalyse, Dissertationsdruck, Nürnberg 1968

Vgl. ders.: Aussagenanalyse. Die Messung von Einstellungen im Text der Aussagen von Massenmedien, unveröffentlichter Forschungsbericht, Nürnberg 1969

Vgl. BLEDJIAN, F.: Ergebnisse und neuere theoretische Ansätze über die Wirkungen der Massenkommunikation auf die Einstellungen der Rezipienten, Dissertationsdruck, Nürnberg 1969

Vgl. STÖSPERG, K.: Die Selbsteinschätzung der Rezipienten und die Kommunikationsdiskrepanz als Determinanten des Einstellungswandels im Massenkommunikationsprozeß, in Vorbereitung.

1 Einleitung

The mathematical theorems, and the numbers, are innumerable, while the days of the researcher are numbered and the practical problems encountered in measurement great.

G. Goude

Wenn man sich den gegenwärtigen Stand der empirischen Forschung und der Theoriebildung in den Sozialwissenschaften vor Augen führt, so scheinen vor allem zwei eng miteinander verknüpfte Probleme Hauptgegenstand der Diskussion zu sein: einmal die Suche nach verhaltensdeterminierenden Variablen und zum anderen die Messung dieser Variablen. Was das erste der beiden Probleme betrifft, so geht es dabei vor allem um Fragen der Theoriebildung, ganz gleich, ob man nun zum gegenwärtigen Zeitpunkt als das Ziel dieser Theoriebildung eine sogenannte Theorie mittlerer Reichweite¹⁾ oder eine umfassende Theorie des menschlichen Verhaltens²⁾ anstrebt.

Die Diskussion von Meßproblemen und vor allem spezieller Meßtechniken gehört eher in den Bereich der Hypothesenprüfung und bezieht sich auf die Verbindung zwischen Begriffen der theoretischen Sprache mit solchen der Beobachtungssprache. Im zweiten und dritten Kapitel dieser Arbeit soll versucht werden, die grundlegenden Probleme der Messung bei der Prüfung sozialwissenschaftlicher Hypothesen aufzuzeigen. Mit dem Konzept der Einstellung scheint ein

1) Vgl. MERTON, R. K.: Social Theory and Social Structure, Glencoe 1957, S. 9, 280, 328

2) Vgl. z.B. das jüngst erschienene Buch von MALEWSKI, A.: Verhalten und Interaktion, Tübingen 1967

Begriff der theoretischen Sprache¹⁾ vorzuliegen, der im Vergleich mit anderen konkurrierenden Begriffen dieser Sprache relativ leicht operationalisierbar ist. Die Brauchbarkeit des Einstellungsbegriffes wird anhand eines theoretischen Ansatzes zur Erklärung von Massenkommunikationsprozessen kurz erläutert: Eine Prüfung der Hypothesen dieses Ansatzes erfordert die Messung der Einstellungen von Kommunikatoren und Rezipienten sowie der im Text von Aussagen der Massenkommunikation geäußerten Einstellungen.

Um die Problematik der Messung zu verdeutlichen, muß zu Anfang die formale Struktur der in den Sozialwissenschaften angestrebten Hypothesen erläutert werden. Nach der hier zugrunde gelegten wissenschaftstheoretischen Auffassung muß es das Ziel sozialwissenschaftlicher Theoriebildung sein, Erklärung und Vorhersage menschlichen Verhaltens zu versuchen.²⁾ Zu diesem Zwecke müssen falsifizierbare Hypothesen aufgestellt werden, die bei einer anschließenden Prüfung strengen Falsifikationsversuchen zu unterziehen sind.

Sodann ist auf den derzeitigen Stand der Theoriebildung und -prüfung in den Sozialwissenschaften näher einzugehen. Dabei wird sich zeigen, daß das zentrale Problem bei der Prüfung dieser Theorien darin besteht, eine Verbindung zwischen den bei der Theoriebildung verwendeten Konstrukten und den Begriffen der Beobachtungsebene herzustellen. Das bedeutet, daß in der Messung der theoretischen Konstrukte die Hauptschwierigkeit bei der Hypothesenprüfung liegt, wohingegen die Prüfung einer Theorie

-
- 1) Vgl. die Diskussion bei ROTH, E.: Einstellung als Determination individuellen Verhaltens, Göttingen 1967
 - 2) Vgl. POPPER, K. R.: Die Zielsetzung der Erfahrungswissenschaft, in: H. ALBERT (Hrsg.): Theorie und Realität, Tübingen 1964, S. 73 - 86

auf logische Konsistenz weit leichter durchzuführen sein dürfte. Nach der Darstellung des Rahmenprogrammes zur Massenkommunikationsforschung, aus dem sich die Zweckmäßigkeit einer Messung von Kommunikator- und Rezipienteneinstellungen ergibt, werden die Grundlagen des Messens in den Sozialwissenschaften unter besonderer Berücksichtigung der Einstellungsmessung behandelt.

Die dabei vertretene Auffassung führt zur Ablehnung eines Operationalismus, der sich damit begnügt, Meßinstrumente zu konstruieren, ohne zu wissen, was diese Instrumente eigentlich messen. Eine solche Auffassung würde die Bildung von Theorien unmöglich machen, da sie auf die Festlegung einer konstitutiven Bedeutung der verwendeten Termini verzichtet. Ebenso führt die vorgeschlagene Argumentation zu einer Ablehnung bloßen theoretischen Spekulierens, weil dabei lediglich die konstitutive Bedeutung verwendeter Termini festgelegt werden kann, ohne den Versuch zu unternehmen, eine Verbindung zur Beobachtungsebene herzustellen und damit prinzipiell das Scheitern einer Theorie an der Realität zu ermöglichen.

Im vierten Kapitel der Arbeit sollen dann die verschiedenen Techniken der Messung von Einstellungen erläutert und einer kritischen Prüfung unterzogen werden. Nach dem bekannten Buch von THURSTONE und CHAVE¹⁾ ist in den vergangenen vierzig Jahren eine Flut von Veröffentlichungen zum Problem der Einstellungsmessung erschienen. In der Zeit kurz nach dem zweiten Weltkrieg wuchs die Literatur auf diesem Gebiet vor allem in den USA sehr rasch:²⁾ Eine Vielzahl von Verfahren der Einstellungs-

1) Vgl. THURSTONE, L. L. und E. J. CHAVE: The Measurement of Attitude, Chicago 1929

2) Vgl. vor allem das Werk von STOUFFER, S.A. u.a.: Measurement and Prediction, Princeton, N. J. 1950, Paperback Ausgabe 1966

messung wurde entwickelt: Zu diskutieren sind die "Judgment"-Techniken, an deren Entwicklung THURSTONE maßgeblich beteiligt war.¹⁾ Es folgen die Techniken der summierten Einschätzungen (LIKERT)²⁾, die Skalogramm-Analyse (GUTTMAN)³⁾, die Diskriminationstechnik (EDWARDS und KILPATRICK)⁴⁾, das Verfahren der transferierten Einschätzungen (COOMBS)⁵⁾ sowie die Analyse latenter Strukturen (LAZARSFELD)⁶⁾.

Damit sind für die Behandlung in der vorliegenden Arbeit die Techniken ausgewählt worden, bei denen die direkte Absicht einer Skalierung von Items und/oder Versuchspersonen besteht, so daß z.B. projektive Tests und ähnliche Vorgehensweisen ebenso ausgeklammert werden können wie die Techniken des semantischen Differentials⁷⁾ und des

-
- 1) Vgl. u.a. die zuletzt erschienene Aufsatzsammlung von THURSTONE, L. L.: The Measurement of Values, Chicago 1959
 - 2) Vgl. LIKERT, R.: A Technique for the Measurement of Attitudes, Arch. Psychol. Nr. 140, 1932, S. 1 - 55
 - 3) Vgl. GUTTMAN, L.: The Quantification of a Class of Attributes: A Theory and Method for Scale Construction, in: P. HORST u.a.: The Prediction of Personal Adjustment, New York 1941, S. 319 - 348
 - 4) Vgl. EDWARDS, A. L. und F. P. KILPATRICK: A Technique for the Construction of Attitude Scales, J. Appl. Psychol. 32, 1948, S. 374 - 384
Vgl. dies.: Scale Analysis and the Measurement of Social Attitudes, Psychometrika 13, 1948, S. 99 - 114
 - 5) Vgl. COOMBS, C. H.: Psychological Scaling without a Unit of Measurement, Psychol. Rev. 57, 1950, S. 145 - 158
 - 6) Vgl. LAZARSFELD, P. F.: The Logical and Mathematical Foundation of Latent Structure Analysis, in: S. A. STOUFFER u.a.: Measurement and Prediction, a.a.O., S. 362 - 412
 - 7) Vgl. OSGOOD, C. E.; G. J. SUCI und P. H. TANNENBAUM: The Measurement of Meaning, Urbana, Ill., 1957

Q-Sorts¹⁾, die nicht unmittelbar zur Konstruktion von Einstellungsskalen führen. Bei der Erläuterung der Techniken kommt es nicht so sehr auf eine detaillierte Ausarbeitung der praktischen Vorgehensweisen mit allen möglichen Variationen, sondern vor allem auf eine Würdigung der jeweils zugrunde liegenden Modelle an. Denn nur eine genaue Kenntnis dieser Modelle ermöglicht eine Beurteilung dessen, was mit den einzelnen Meßverfahren erreicht werden kann.

Das letzte Kapitel der Arbeit schließlich enthält einige Bemerkungen zum gegenwärtigen Stand der Konstruktion von Einstellungsskalen.

1) Vgl. STEPHENSON, W.: The Study of Behavior, Chicago und London 1953

2 Theoriebildung und Messung

21 Das Ziel der empirischen Wissenschaften

Das grundlegende Anliegen der Natur- und der Sozialwissenschaften besteht in der *B e s c h r e i b u n g* empirisch auffindbarer Phänomene einerseits und in der *E r k l ä r u n g* und Vorhersage solcher Phänomene mit Hilfe von Hypothesen und Theorien andererseits. Man kann - der Argumentation TORGERSON¹⁾ und MARGENAU²⁾ folgend - die verschiedenen wissenschaftlichen Disziplinen danach unterscheiden, wieweit es ihnen bisher gelungen ist, Theorien mit Erklärungskraft zu entwickeln: Auf der einen Seite stehen die Disziplinen, die sich bisher mit der Beschreibung der Beziehungen zwischen mehr oder weniger direkt beobachtbaren Variablen begnügten oder begnügen mußten (*correlational procedures*), auf der anderen Seite die Disziplinen, die versuchten, diese beobachteten Beziehungen mit Hilfe von Hypothesen und Theorien zu erklären (*theoretical procedures*).

Unter einer Erklärung soll in diesem Zusammenhang verstanden werden: "eine Klasse von Sätzen ..., von denen einer den Sachverhalt beschreibt, der erklärt werden soll (das *e x p l i c a n d u m*), während die anderen die erklärenden Aussagen, die 'Erklärung' im engeren Sinne des Wortes bilden (das *e x p l i c a n s* des *e x p l i c a n d u m s*)"³⁾. Um befriedigende Erklärungen zu finden, wird eine Theorie benötigt, das heißt "eine Menge (System) durch Ableitbarkeitsbeziehungen miteinander verbundener nomologi-

1) Vgl. TORGERSON, W. S.: *Theory and Methods of Scaling*, New York, London und Sydney 1953, S. 1 f.

2) Vgl. MARGENAU, H.: *The Nature of Physical Reality*, New York, Toronto und London 1950, S. 27 - 30

3) POPPER, K. R.: *Die Zielsetzung der Erfahrungswissenschaft*, a.a.O., S. 73

scher Hypothesen"1), wobei eine Hypothese eine behauptete Beziehung zwischen mindestens einer unabhängigen Variablen (Determinante) und mindestens einer abhängigen Variablen (Resultante) ist. Nomologische Hypothesen sind solche, deren Geltungsbereich von Raum und Zeit unabhängig ist.

Die Sozialwissenschaften und so auch die Disziplin, die sich mit den Problemen der Massenkommunikation beschäftigt, gehören bisher im wesentlichen noch zu den beschreibenden Wissenschaften. Ziel muß es jedoch sein, auch die Sozialwissenschaften zu einer theoretischen Wissenschaft im eingangs erläuterten Sinne weiterzuentwickeln. So ist der Satz POPPERs zu verstehen: "Ich nehme an, daß es das Ziel der empirischen Wissenschaft ist, b e f r i e d i g e n d e E r k l ä r u n g e n zu finden für alles, was uns einer Erklärung zu bedürfen scheint"2). Voraussetzung für die Entwicklung zu einer theoretischen Wissenschaft ist die adäquate Messung der in den Hypothesen enthaltenen Variablen.

Im Rahmen der vorliegenden Arbeit werden vorwiegend Fragen der Syntax und der Semantik am Problem der Messung von Einstellungen erörtert. Während sich die Syntax mit den Beziehungen zwischen den Symbolen, die in einem Modell Verwendung finden können, beschäftigt und dabei von konkret beobachtbaren Objekten abstrahiert, behandelt die Semantik die Beziehungen zwischen Symbolen und Objekten bzw. Eigen-

1) ALBERT, H.: Probleme der Theoriebildung, in: H. ALBERT (Hrsg.): Theorie und Realität, a.a.O., S. 27

2) POPPER, K. R.: Die Zielsetzung der Erfahrungswissenschaft, a.a.O., S. 73

schaften von Objekten.¹⁾ Mit Hilfe dieser Unterscheidung kann bereits ein Hauptproblem der Arbeit erstmals genauer charakterisiert werden: Innerhalb eines (uninterpretierten) syntaktischen Systems kann über die Richtigkeit eines Satzes eine endgültige Aussage getroffen werden. Sätze sind richtig, wenn sie den festgelegten Regeln des betreffenden Systems entsprechen. "But these absolutely true statements say nothing at all, unfortunately, about objects or events, and they reveal no secrets of nature".²⁾ Durch die Regeln der Semantik, die z.B. Instrumente zur Messung von Einstellungen sein können, wird ein syntaktisches System (z.B. ein mathematisches Modell) mit der Beobachtungsebene verbunden und kann auf diese Weise auf seine Adäquanz für ein bestimmtes empirisches Phänomen geprüft werden. Eine so erfolgende Interpretation eines syntaktischen Systems muß immer unvollständig bleiben. Die Gründe dafür sind sehr verschieden: Die Meßinstrumente - als eine Art von semantischen Regeln - können zu fehlerhaften Ergebnissen führen, bestimmte Termini des syntaktischen Systems bleiben uninterpretiert oder sind überhaupt uninterpretierbar,³⁾ zudem werden nur einige der Aspekte eines Phänomens von dem Modell erfaßt.

-
- 1) Vgl. hierzu und zu den folgenden Überlegungen: STEVENS, S. S.: Mathematics, Measurement, and Psychophysics, in: S. S. STEVENS (Hrsg.): Handbook of Experimental Psychology, New York, London und Sydney 1951, S. 2 ff. Von den Problemen der "pragmatics", die sich auf die Beziehungen zwischen Symbolen und den Benutzern dieser Symbole beziehen, wird an einigen Stellen dieser Arbeit ebenfalls die Rede sein.
- 2) STEVENS, S. S.: Mathematics, Measurement, and Psychophysics, a.a.O., S. 3
- 3) STEVENS bringt das Beispiel der Hypotenuse im rechtwinkligen Dreieck, deren Länge sich aus der nach c aufgelösten Formel $c^2 = a^2 + b^2$ berechnet. Eine mögliche Auflösung ist $c = -\sqrt{a^2 + b^2}$; diese Lösung wird jedoch als uninterpretierbar ausgeschieden, weil sie "keinen Sinn ergibt".

Die wissenschaftstheoretische Diskussion hat sich in der Vergangenheit vorwiegend mit Problemen der Syntax theoretischer Systeme beschäftigt und dabei die Fragen der Semantik oftmals vernachlässigt. Ein Grund für diesen Sachverhalt kann darin liegen, daß die moderne Wissenschaftstheorie durch verschiedene philosophische Strömungen beeinflusst ist, bei denen Meßprobleme als weniger oder nicht relevant abgetan wurden. Andererseits mangelte es denjenigen, die sich z.B. in der Psychologie intensiv mit der Konstruktion von Meßinstrumenten beschäftigen, oftmals an ausreichenden wissenschaftstheoretischen Kenntnissen. Aber seitdem die Forderung nach Aufstellung empirisch prüfbarer Theorien und nach empirischer Forschung in den Sozialwissenschaften allgemein akzeptiert wird, ist noch relativ wenig zur Klärung des Verhältnisses von Theoriensprache und Beobachtungssprache geschehen. Und hierin ist sicherlich eine Ursache dafür zu sehen, daß die Sozialwissenschaften bisher im Vergleich zu den Naturwissenschaften nur wenig leistungsfähige Theorien vorweisen können.

Bei den Ausführungen in diesem und dem folgenden Kapitel geht es um eine Erörterung der Grundlagen der Messung, die im Zusammenhang mit dem Problem der Theoriebildung diskutiert werden sollen. Die dabei auftretenden Fragen lauten: Welche Rolle spielt die Messung von Variablen bei der Hypothesenbildung und -prüfung? Was ist überhaupt unter Messung zu verstehen? Welche Anforderungen sind an ein Meßinstrument zu stellen, damit es bei der Prüfung von Hypothesen eingesetzt werden kann? - Um einen Beitrag zur Klärung dieser Fragen leisten zu können, soll zunächst die Struktur einer voll entwickelten Theorie, wie sie etwa in den Naturwissenschaften besteht, erläutert werden, um anschließend einen Vergleich mit den bisherigen theoretischen Bemühungen in den Sozialwissenschaften anstellen zu können.

22 Die formale Struktur voll entwickelter Theorien

Wie bereits angedeutet, besteht die im Rahmen einer empirischen Wissenschaft auszuführende Tätigkeit in der Bildung von Theorien und in der Sammlung von Daten.¹⁾ Die formale Struktur einer Theorie kann anhand der verwendeten theoretischen Konstrukte²⁾ und deren Beziehungen zueinander dargestellt werden. Die empirische Seite besteht aus den beobachtbaren Daten. Durch Korrespondenzregeln³⁾ werden beide miteinander verbunden. Das bedeutet, daß bestimmte theoretische Konstrukte ganz oder teilweise in den Formen beobachtbarer Daten definiert werden. Zwischen den einzelnen theoretischen Konstrukten bestehen logische formale Beziehungen.

Für ein System theoretischer Konstrukte, das durch logische formale Beziehungen verbunden ist, verwendet TOSERSON den Terminus "Modell". Wenn zumindest ein Teil der Konstrukte durch Korrespondenzregeln mit der Beobachtungsebene verknüpft ist, wird aus dem Modell eine Theorie, die einer empirischen Prüfung unterzogen werden kann. Ein Modell, in dem kein Konstrukt empirisch interpretierbar ist, kann nur auf logische Konsistenz, nicht aber auf Obereinstimmung mit der Realität geprüft werden.

- 1) Die Darstellung dieses Abschnittes geht zunächst aus von TOSERSON, H. G.: *Theory and Methods of Scaling*, a.a.O., S. 2 - 5, der seinerseits auf die Ausführungen MARGENAU aufbaut, vgl. MARGENAU, H.: *The Nature of Physical Reality*, a.a.O., S. 236 et passim
- 2) Synonyme zum Begriff des theoretischen Konstruktes stellen die Ausdrücke "Dispositionstermini", "intervenierende Variable", "hypothetische Eigenschaften" und "theoretische Terme" dar. In dieser Arbeit wird der Ausdruck Konstrukt bzw. theoretisches Konstrukt verwendet, da er am wenigsten Anlaß zu Mißverständnissen geben dürfte. Der diesbezügliche Vorteil des hier gewählten Begriffes wird besonders deutlich, wenn man ihn mit dem in mancherlei Bedeutung gebräuchlichen Begriff der intervenierenden Variablen vergleicht.
- 3) Die Termini "operationale Definition" und "Interpretationsregel" können als Synonyme zum Begriff der Korrespondenzregel betrachtet werden.

Daraus geht hervor, daß die in einer Theorie verwendeten Konstrukte auf zweierlei Weise definiert sein können: einmal in den Termini anderer Konstrukte. So ist z.B. eine formale Gleichung, die die Beziehung zwischen zwei oder mehr Konstrukten ausdrückt, als Definition jedes in der betreffenden Gleichung vorkommenden Konstruktes in den Termini der anderen Konstrukte anzusehen. In diesem Sinne spricht man von einer konstitutiven Definition¹⁾, die dem Konstrukt theoretische oder systematische Bedeutung verleiht.²⁾ TORGERSON bildet hierfür ein Beispiel aus dem naturwissenschaftlichen Bereich anhand der Konstrukte Kraft, Masse und Beschleunigung³⁾: Kraft ist das Produkt aus Masse und Beschleunigung, Masse ist das Verhältnis von Kraft zu Beschleunigung, und Beschleunigung kann definiert werden als das Verhältnis von Masse zu Kraft. Die zweite mögliche Art einer Definition von Konstrukten erfolgt durch die Korrespondenzregeln, die ihnen operationale Bedeutung verleihen. Unter Korrespondenzregel soll hier die Gesamtheit aller Anweisungen zur Messung eines bestimmten Konstruktes angesehen werden. (Mit solchen Regeln zur Messung des theoretischen Konstruktes "Einstellung" beschäftigt sich das Kapitel 4 dieser Arbeit.)

Alle in einer Theorie verwendeten Konstrukte müssen konstitutive Bedeutung besitzen. Dagegen brauchen nicht alle Konstrukte auf die zweite Weise, nämlich in Form von Korrespondenzregeln, in den Termini beobachtbarer Daten definiert zu sein. Konstrukte ohne solche Korrespondenzregeln müssen jedoch indirekt auf dem Wege über andere Konstrukte mit

1) Vgl. MARGENAU, H.: The Nature of Physical Reality, a.a.O., S. 236

2) Vgl. HEMPEL, C. G.: Fundamentals of Concept Formation in Empirical Science, in: O. NEURATH u.a. (Hrsg.): International Encyclopedia of Unified Science, Band 2, Nr. 7, Chicago, London und Toronto 1952, S. 39

3) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 4

der Beobachtungsebene verbunden sein, um in einer Theorie, die erklären soll, verwendet werden zu können. Ein solches Konstrukt ohne Korrespondenzregeln ist in der physikalischen Theorie z.B. die Dichte als Verhältnis von Masse zum Volumen einer Substanz: Das Konstrukt "Dichte" selbst braucht zwar keine Korrespondenzregeln zu haben, muß jedoch über die Konstrukte "Masse" und "Volumen" mit der Beobachtungsebene verbunden sein, wobei natürlich die beiden Konstrukte "Masse" und "Volumen" gegebenenfalls auch wieder lediglich auf dem Wege über andere Konstrukte mit der Beobachtungsebene in Verbindung stehen könnten.

Ein theoretisches System mit der geschilderten Struktur hat den Vorteil, daß man, von bestimmten beobachtbaren Daten ausgehend, mit Hilfe einer Korrespondenzregel in den theoretischen Raum gelangen und dort z.B. mittels einer mathematischen Gleichung von einem Konstrukt zu einem oder mehreren anderen Konstrukten fortschreiten kann, um schließlich aufgrund einer zweiten Korrespondenzregel zur Beobachtungsebene zurückzukehren.¹⁾ Durch einen Vergleich zwischen einem auf diese Weise vorhersagbaren Meßergebnis bei der Anwendung der zweiten Korrespondenzregel und dem tatsächlichen Meßergebnis ist es möglich, ein eventuell umfassendes theoretisches System durch relativ wenige empirische Untersuchungen auf Übereinstimmung mit der Wirklichkeit zu testen.

Bei den bisherigen Ausführungen dieses Abschnittes wurde weitgehend die Terminologie TORGERSONs verwendet. Es muß jedoch angemerkt werden, daß diese Terminologie im Gebrauch des Begriffes "Definition" mißverständlich ist, weil

1) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 3 f.

dieser Begriff hier nicht im üblichen Sinne der klassischen Logik verwendet wird. Was TORGERSON mit "Definition" bezeichnet, wird treffender und weniger mißverständlich durch den Terminus "Postulat" ausgedrückt. In diesem Sinne stellt CARNAP fest, daß beim nichtlogischen Vokabular einer Wissenschaftssprache zwei Arten von Termini unterschieden werden können: "die Beobachtungsterme (z.B. 'blau') und die theoretischen Terme (z.B. 'elektrisches Feld'). Die letzteren werden nicht durch Definitionen (im klassischen Sinne¹⁾) eingeführt, sondern durch Postulate zweier Arten, nämlich theoretische Postulate, z.B. Grundgesetze der Physik, und Korrespondenzpostulate, die die theoretischen Terme mit Beobachtungstermini verbinden"²⁾. Wenn TORGERSON von "Modell" spricht, meint er dasselbe wie CARNAP mit dem Ausdruck "ungedeutetes Kalkül": "a system of formal rules which determine certain formal properties and relations of sentences, especially for the purpose of formal deduction"³⁾. Als Beispiel für ein solches Modell (oder ungedeutetes Kalkül) kann die theoretische Physik angesehen werden. Die in einem solchen Modell verwendeten "theoretischen Terme erhalten durch die Korrespondenzpostulate wenigstens eine unvollständige Deutung"⁴⁾.

Damit ist die Frage gestellt, wie weit die Konstrukte einer Theorie in ihrer konstitutiven und in ihrer operationalen Bedeutung übereinstimmen: Es besteht eine Lücke zwischen der Beobachtungssprache und der theoretischen Sprache. Dieses Problem soll im folgenden Abschnitt anhand der bisherigen Versuche der Theoriebildung in den Sozialwissen-

1) Anmerkung des Verfassers

2) CARNAP, R.: Beobachtungssprache und theoretische Sprache, in: Dialektika 12, 1958, S. 248
An anderer Stelle verwendet jedoch auch CARNAP den Begriff Definition für das hier Gemeinte.

3) CARNAP, R.: Elementary and Abstract Terms, in: A. DANTO und S. MORGENBESSER (Hrsg.): Philosophy of Science, New York 1960, S. 151 (Abdruck der Abschnitte 24 und 25 aus R. CARNAP: Foundations of Logic and Mathematics, Encyclopedia of Unified Science, Bd. 3, Chicago 1937)

4) CARNAP, R.: Beobachtungssprache und theoretische Sprache, a.a.O., S. 248

schaften näher erläutert werden. An dieser Stelle sei lediglich noch eine Bemerkung zu der Vorgehensweise beim Aufbau von Theorien der geschilderten Struktur eingefügt. Grundsätzlich sind zwei Vorgehensweisen denkbar¹⁾: Man kann entweder mit der Einführung von Konstrukten nahe an der Beobachtungsebene beginnen, um von da aufbauend weitere Konstrukte höheren Abstraktionsgrades einzuführen. Eine solche Vorgehensweise erleichtert die Prüfung der Frage, wie ein bestimmtes Konstrukt empirisch verankert ist. Wie CARNAP feststellt, zeigt jedoch die Geschichte der Wissenschaft, daß es auf diese Weise bisher nicht gelungen ist, zu leistungsfähigen Theorien zu gelangen. Es handelt sich dabei nach CARNAPs Ansicht um eine empirische Tatsache, nicht aber um eine logische Notwendigkeit.²⁾ Bei der zweiten Vorgehensweise beginnt man gewissermaßen an der Spitze des Systems, um dann auf immer niedrigere Abstraktionsebenen abzustiegen: "It consists in taking a few abstract terms as primitive signs and a few fundamental laws of great generality as axioms. Then further terms, less and less abstract, and finally elementary ones, are to be introduced by definitions."³⁾ Wenn auch CARNAP diesem zweiten Weg den Vorzug gibt, so ist damit doch die Schwierigkeit verbunden, daß in diesem Fall die Übereinstimmung eines eingeführten Konstruktes in seiner konstitutiven und in seiner operationalen Bedeutung besonders problematisch ist. Zudem dürften im Laufe der Geschichte der Wissenschaft beide Wege angewendet worden sein, wenn auch mit fortschreitendem Ausbau des theoretischen Gebäudes einer Wissenschaft der zweite Weg zunehmend beschritten zu werden scheint.

1) Vgl. CARNAP, R.: Elementary and Abstract Terms, a.a.O., S. 152 ff.

2) Die Problematik dieser Behauptung soll hier nicht erörtert werden.

3) CARNAP, R.: Elementary and Abstract Terms, a.a.O., S. 153

23 Zum Stand der Theoriebildung und -prüfung in den Sozialwissenschaften

Die Vorgänge des Messens bzw. der Theoriebildung und -prüfung sind in den Sozialwissenschaften grundsätzlich dieselben wie in den Naturwissenschaften. In den Sozialwissenschaften bestehen jedoch spezifische "Schwierigkeiten bei der Anwendung quantitativer Methoden, insbesondere der Meßmethoden".¹⁾ POPPER sieht hierin einen der Hauptunterschiede "zwischen den Methoden einiger der theoretischen Wissenschaften von der Natur und von der Gesellschaft"²⁾, wobei es sich nach seiner Ansicht aber "um graduelle, nicht um essentielle Differenzen handelt"³⁾. Wenn solche spezifischen Schwierigkeiten in den Sozialwissenschaften bestehen, so bedeutet das in den Ausdrücken der hier verwendeten Terminologie u.a., daß es in den Sozialwissenschaften bisher weniger gut als in den Naturwissenschaften gelungen ist, die Lücke zwischen Beobachtungssprache und Theoriesprache zu schließen. Ein vor kurzer Zeit in einem Buch zur Methodologie sozialwissenschaftlicher Forschung erschienener Aufsatz trägt den bezeichnenden Titel: "The Measurement Problem: A Gap between the Languages of Theory and Research"⁴⁾. Dieser Aufsatz wird in einem Vorwort zum ersten Teil des Buches, der sich mit den Problemen von Messung und Begriffsbildung beschäftigt, mit folgenden Bemerkungen eingeführt: "The central thesis of this chapter is

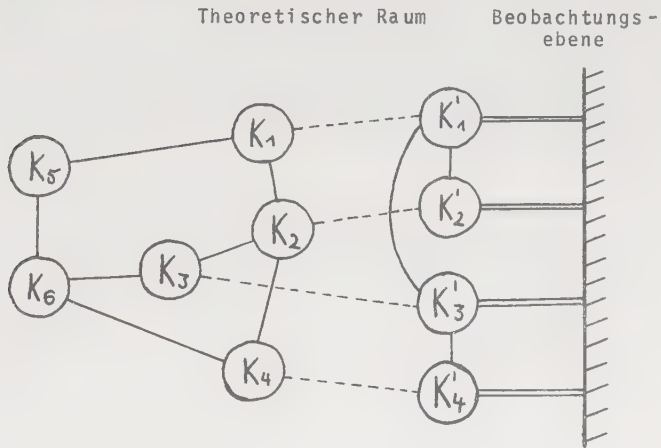
1) POPPER, K. R.: Das Elend des Historizismus, Tübingen 1965, S. 111

2) Ebenda

3) ders., a.a.O., S. 110

4) BLALOCK, H. M. Jr.: The Measurement Problem; A Gap between the Languages of Theory and Research, in: H. M. BLALOCK Jr. und A. B. BLALOCK (Hrsg.): Methodology in Social Research, New York, St. Louis, San Francisco, Toronto, London und Sidney 1965, S. 5 - 27

that measurement process is an indirect one, there are no purely logical or deductive ways of establishing linkages between theoretically defined variables and actual measures, and that under some circumstances theory and measurement become intertwined in a rather complex manner¹⁾. Die formale Struktur sozialwissenschaftlicher Theorien läßt sich anhand der folgenden schematischen Darstellung veranschaulichen:



Figur 1: Schematische Darstellung der möglichen Formalstruktur einer sozialwissenschaftlichen Theorie²⁾

1) BLALOCK, H. M. Jr. und A.B.BLALOCK (Hrsg.): Methodology in Social Research, a.a.O., S. 2; Vorwort zum Teil 1: "Measurement and Conceptualization"

2) Die Darstellung wurde leicht verändert übernommen von TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 5, Figur 2. Es erscheint zweifelhaft, ob in den Sozialwissenschaften bereits Theorien mit einer so komplizierten Struktur wie in diesem Schema entwickelt wurden. Selbst wenn das nicht der Fall ist, kann die Darstellung dennoch zur Erläuterung der Problematik nützlich sein.

Die Darstellung zeigt zwei Arten von Konstrukten: Die einen (K'_1 bis K'_4) liegen nahe an der Beobachtungsebene, sind untereinander verbunden (einfache Linien) und stehen durch Korrespondenzregeln (doppelte Linien) mit der Beobachtungsebene in Verbindung. Die andere Art von Konstrukten (K_1 bis K_6) ist wohl untereinander verbunden (einfache Linien) besitzt aber keine Korrespondenzregeln. Eine zunächst unklare Verbindung zwischen den beiden Arten von Konstrukten soll im Schema durch die gestrichelten Linien ausgedrückt werden.

Bei den bisherigen theoretischen Versuchen in den Sozialwissenschaften bestehen nun im Vergleich zum Idealfall einer voll entwickelten Theorie vor allem zwei erhebliche Mängel, die mit Hilfe der schematischen Darstellung erläutert werden können¹⁾:

- a) Was die Konstrukte K_1 bis K_6 betrifft, so werden die Beziehungen zwischen ihnen in aller Regel nicht in Form von Gleichungen mathematisch ausgedrückt wie etwa in der Physik oder der Chemie, sondern sie werden verbal und damit wohl auch weniger präzise formuliert. Beispiel eines gescheiterten Versuches in dieser Hinsicht auf dem Gebiet der Einstellungsforschung Besserung zu schaffen, bildet eine Arbeit von TANNENBAUM²⁾, in der das Kongruenzmodell³⁾ einer experimentellen Prüfung unterzogen wurde. In seiner Würdigung dieser Arbeit bemerkt BROWN: "The general assumptions of the congruity model have been confirmed by experiment, but the arithmetical details have not."⁴⁾

- 1) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., vor allem S. 7 f.
- 2) Vgl. TANNENBAUM, P. H.: Attitudes Toward Source and Concept as Factors in Attitude Change through Communications, Unveröffentlichte Dissertation, Illinois 1953
- 3) Vgl. BLEDDJIAN, F.: Ergebnisse und neuere theoretische Ansätze über die Wirkungen der Massenkommunikation auf die Einstellungen der Rezipienten, a.a.O., S. 154 ff. Siehe hier auch weitere Literatur zum Kongruenzmodell.
- 4) BROWN, R.: Models of Attitude Change, in: New Directions in Psychology I, New York, Chicago, San Francisco, Toronto und London 1962, S. 34

- b) Die Konstrukte K'_1 bis K'_4 haben zwar Korrespondenzregeln, so daß Meßoperationen prinzipiell ebenso möglich sind wie die Bestimmung der Beziehungen zwischen ihnen mittels Korrelationskoeffizienten oder ähnlichen mathematischen Ausdrücken. Aber: Die Entscheidung darüber, ob die verwendeten Konstrukte in ihrer operationalen Bedeutung (K') und in ihrer konstitutiven Bedeutung (K) übereinstimmen, ist eine Frage empirischer Prüfung. In den Sozialwissenschaften stehen theoretisches Konstrukt und Meßvorgang oft relativ unabhängig nebeneinander. Die operational definierten Konstrukte (K') sind Indikator oder Index dessen, was man für ihre konstitutiv definierten Äquivalente (K) hält. Diese Indikatoren "are effects or correlates related to psychological dimensions by unknown laws".¹⁾

Das kann an folgendem Beispiel illustriert werden: Ein Test mißt ein bestimmtes Konstrukt K' , das in der Sprache der Psychologie "Intelligenz" genannt wird. Das Konstrukt K' ("Intelligenz") hat damit unmittelbare operationale Bedeutung. Gleichzeitig bestehe nun eine Theorie oder zumindest eine Hypothese, in der ein Konstrukt K , das ebenfalls "Intelligenz" genannt wird, in einer bestimmten konstitutiven Bedeutung in den Termini anderer Konstrukte festgelegt ist. Es ist nun keineswegs klar, daß das Konstrukt "Intelligenz" in seiner konstitutiven und in seiner operationalen Bedeutung übereinstimmt. Grundsätzlich verhält es sich mit dem hier zu behandelnden Konstrukt "Einstellung" ebenso: Die Einstellung des Rezipienten gegenüber der Zeitung ist beispielsweise eine Variable, die als theoretisches Konstrukt K

1) STEVENS, S. S.: Mathematics, Measurement, and Psychophysics, a.a.O., S. 47

in eine Hypothese eingebaut sein kann. Nun gibt es eine Skala, mit der diese Variable "Einstellung gegenüber der Zeitung" gemessen werden soll.¹⁾ Nichts kann jedoch a priori darüber gesagt werden, wie weit das Konstrukt "Einstellung gegenüber der Zeitung" in seiner operationalen Bedeutung K' und in seiner konstitutiven Bedeutung K übereinstimmt. Die Situation wird besonders dadurch kompliziert, daß für K und für K' dieselben Namen verwendet werden, obwohl es sich eventuell um zwei völlig verschiedene Konstrukte handeln kann. Hinzu kommt noch, daß die K-Konstrukte in den Sozialwissenschaften oftmals sehr stark mit einer vorwissenschaftlichen und daher ungenauen Bedeutung belastet sind²⁾.

Damit ist das Problem angesprochen, das allgemein in den Lehrbüchern über Forschungstechnik unter dem Stichwort "Validität" ("Gültigkeit") behandelt wird.³⁾ Die am weitesten verbreitete Definition von Validität lautet: Ausmaß, in dem ein Meßinstrument das mißt, was es zu messen vorgibt. Die Frage nach der Validität eines Meßinstrumentes kann, wie gesagt, nur durch empirische Prüfung geklärt werden.⁴⁾

Besonders in den Sozialwissenschaften besteht bisher die Schwierigkeit, daß man über die Art der Verbindung zwischen den K'-Konstrukten und den K-Konstrukten, auf deren

1) Vgl. ROGERS, J. L.: Prospective Teacher's Attitudes Toward Freedom of Information, in: Journalism Quarterly 32, 1955, S. 169 - 176

2) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 8

3) Vgl. z.B. SELLITZ, C., M. JAHODA, M. DEUTSCH und S. W. COOK: Research Methods in Social Relations, New York, Chicago, San Francisco und Toronto 1959, S. 154 ff.

4) Die Problematik der Vorgehensweisen bei empirischen Validitätsprüfungen wird im Abschnitt 3322 behandelt.

Messung es im Rahmen der Konstruktion leistungsfähiger Theorien eigentlich ankommt, oftmals keine oder doch nur ungenügende Kenntnisse hat.¹⁾ KAPLAN nennt die Vorstellung einer validen fehlerfreien Messung "the fiction of the true measure"²⁾. Man kann lediglich versuchen, Evidenz für die Gültigkeit eines Instrumentes zu sammeln, um als nicht valide erkannte Meßinstrumente auszuschneiden oder eine entsprechende Modifizierung der theoretischen Konstrukte vorzunehmen. Es sei jedoch nochmals betont, daß hierin kein grundsätzlicher Unterschied zu den Naturwissenschaften gesehen werden kann; nur ist es in den Naturwissenschaften besser als in den Sozialwissenschaften gelungen, die Beziehungen zwischen den K'- und den K-Konstrukten zu klären.³⁾

Aus der unsicheren Beziehung zwischen konstitutiver und operationaler Bedeutung eines Konstruktes ergeben sich einige sehr wichtige Folgerungen für die Hypothesenbildung wie für die Hypothesenprüfung. Es besteht folgendes Dilemma:⁴⁾

-
- 1) Die Unsicherheit dieser Beziehungen soll im Schema durch die gestrichelte Linie zum Ausdruck gebracht werden.
 - 2) KAPLAN, A.: The Conduct of Inquiry, San Francisco 1964, S. 202
 - 3) Obwohl sich TORGERSON in dieser Beziehung nicht klar äußert, scheint doch aus seinen Ausführungen hervorzugehen, daß er hier - im Gegensatz zu der in dieser Arbeit vertretenen Auffassung - einen essentiellen Unterschied zwischen Natur- und Sozialwissenschaften sieht. Dieser Eindruck entsteht besonders, wenn man seine schematische Darstellung einer voll entwickelten Theorie mit dem Schema einer für die Sozialwissenschaften typischen Theorie vergleicht.
Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 3 Figur 1 und S. 5 Figur 2
 - 4) Vgl. ders., a.a.O., S. 7 f.

- a) Wenn die Ergebnisse einer Messung mit den Voraussagen einer Hypothese übereinstimmen, ist Evidenz für die Richtigkeit dieser Hypothese gesammelt worden. Stimmt das Meßergebnis jedoch nicht mit der Voraussage der Hypothese überein, so kann sie als falsifiziert abgelehnt werden.
- b) Wenn das Meßergebnis **n i c h t** mit der Voraussage der Hypothese übereinstimmt, kann man sich jedoch auch dafür entscheiden, die operational definierten Konstrukte (K') nicht als Maß für die konstitutiv definierten Konstrukte (K) anzuerkennen. Das hat zur Folge, daß die Hypothese nicht als falsifiziert anzusehen ist, das verwendete Meßinstrument dagegen als ungemessen ausgeschieden werden kann.

Hypothesen über die Beziehungen zwischen Variablen, die als Konstrukte mit konstitutiver Bedeutung formuliert sind (K), erweisen sich so als relativ immun gegenüber widersprechenden Ergebnissen, die in empirischen Untersuchungen gewonnen werden, bei denen die Beziehungen zwischen den entsprechenden operational definierten Konstrukten (K') untersucht werden.¹⁾

Das genannte Dilemma ist grundsätzlich unaufhebbar. Es gibt keine Gültigkeitskriterien: Ebensowenig wie Hypothesen endgültig verifiziert werden können, kann die Gültigkeit eines Meßinstrumentes endgültig nachgewiesen werden; denn mit Hilfe der Verfahren, die eine Prüfung der Frage erlauben, wieweit die in einer Hypothese verwendeten Konstrukte in ihrer konstitutiven und in ihrer operationalen Bedeutung übereinstimmen, lassen sich allenfalls **näherungsweise** Lösungen erzielen. Trotzdem sollte eine der Hauptaufgaben der Forschung in der Suche nach solchen Konstrukten beste-

1) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 7 f.

hen, bei denen die Beziehungen zwischen konstitutiver und operationaler Bedeutung möglichst weitgehend geklärt werden können. Für den Bereich der Massenkommunikationsforschung bietet sich in dieser Beziehung - wie auch in anderen Bereichen sozialwissenschaftlicher Forschung - das Konstrukt "Einstellung" an.¹⁾ Um eine Entscheidung über die Weiterverwendung oder Zurückweisung einer Hypothese oder eines Meßinstrumentes möglichst sinnvoll fällen zu können²⁾, ist die genaue Kenntnis des Meßvorganges und der Beschaffenheit der verwendeten Meßinstrumente ebenso nötig wie die Kenntnis der logischen Grundlagen der Hypothesenbildung und -prüfung. (Mit dem ersten Problem beschäftigen sich die Kapitel 3 und 4 dieser Arbeit.)

An dieser Stelle sei noch eine Bemerkung zu einer im wesentlichen zu dem zweiten Problemkreis gehörenden Frage eingefügt, die im folgenden nicht erneut aufgegriffen wird: Es handelt sich dabei um das Problem der Asymmetrie von Verifizierbarkeit und Falsifizierbarkeit.³⁾ Es wird behauptet, daß eine solche Asymmetrie für theoretische Aussagen in der Hinsicht bestehe, daß diese Aussagen zwar nicht endgültig verifizierbar, aber doch endgültig falsifizierbar seien. An Theorien ist demnach die Forderung der Falsifizierbarkeit zu stellen. Die Ausführungen dieses Kapitels lassen jedoch an der Brauchbarkeit dieses Kriteriums zweifeln: Wenn es unmöglich ist, die Validität eines Meßinstrumentes endgültig nachzuweisen, so ist es auch unmöglich, theoretische Aussagen⁴⁾

1) Vgl. hierzu das nachfolgende Kapitel 24

2) Sinnvoll bedeutet hier: im Hinblick auf den wissenschaftlichen Fortschritt, und das bedeutet im Sinne KUHNs die Erörterung der Frage, "welche von zwei miteinander konkurrieren Theorien besser zu den Fakten paßt."
KUHN, T. S.: Die Struktur wissenschaftlicher Revolutionen, Frankfurt 1967, S. 195 et passim, zuerst unter dem Titel: The Structure of Scientific Revolutions, Chicago 1962

3) Vgl. POPPER, K. R.: Logik der Forschung, 2. deutsche Auflage, Tübingen 1966, S. 14 - 17 et passim

4) Unter theoretischen Aussagen werden hier nomologische Hypothesen verstanden.
Vgl. ALBERT, H.: Probleme der Theoriebildung, a.a.O., S. 23 f.

endgültig zu falsifizieren, weil eine solche Falsifikation die valide Messung der in den Hypothesen verwendeten Konstrukte voraussetzen würde. Denn auch aus nomologischen Hypothesen aufgebaute Theorien weisen die in der Figur 1 dargestellte Struktur auf, so daß das genannte Dilemma auch bei dieser Form der Hypothesen bestehen bleibt: Erbringt die Anwendung eines Instrumentes ein mit der betreffenden Hypothese nicht zu vereinbarendes Meßergebnis, so ist eine Entscheidung zu fällen, ob man mit der Hypothese weiterarbeiten und das Instrument als nicht valide zurückweisen soll (das wird man eventuell tun, wenn die Hypothese Bestandteil einer ansonsten gut geprüften Theorie ist), oder ob man davon ausgeht, genügend Evidenz für die Validität des Instrumentes gesammelt zu haben und dementsprechend die Hypothese als falsifiziert ansieht.¹⁾

24 Die Bedeutung der Einstellungsmessung für die Bildung und Prüfung von Hypothesen zur Erklärung von Massenkommunikationsprozessen

241 Die wichtigsten Vorteile des Einstellungsbegriffes

Wie bereits im Vorwort bemerkt, ist diese Arbeit im Rahmen eines umfangreichen Forschungsprogrammes zu Problemen der Massenkommunikation entstanden. Eine der Voraussetzungen zur Durchführung dieses Programmes ist die re-

1) Selbst wenn man die Hypothese als falsifiziert ansieht, ist es im übrigen immer noch zweifelhaft, ob sie damit auch als Grundlage weiterer Forschungen ausgeschieden wird, solange noch keine brauchbarere Hypothese zur Verfügung steht.
Vgl. KUHN, T. S.: Die Struktur wissenschaftlicher Revolutionen, a.a.O., S. 194 f.

liable und valide Messung der Einstellungen von Kommunikatoren und Rezipienten sowie der Einstellungen, die in den von den Massenmedien verbreiteten Aussagen geäußert werden. Denn der Basisbegriff des theoretischen Konzeptes, das dem Forschungsprojekt zugrunde liegt, ist der Einstellungsbegriff. Mit dem grundlegenden Problem der Messung der Einstellungen von Kommunikatoren und Rezipienten beschäftigt sich die vorliegende Arbeit; die Möglichkeiten der Messung von Einstellungen in den Aussagen der Massenmedien wurden bereits in einer anderen Arbeit untersucht.¹⁾ An dieser Stelle kommt es nun darauf an, die Vorteile des Einstellungsbegriffes sowie den auf diesem Begriff aufbauenden theoretischen Ansatz kurz zu charakterisieren²⁾, bevor dann im folgenden die Grundlagen der Messung und vor allem die einzelnen Möglichkeiten der Einstellungsmessung ausführlich diskutiert werden.

Die weitgehende Unzufriedenheit mit dem gegenwärtigen Stand der Hypothesenbildung und -prüfung über den Prozeß der Massenkommunikation besteht zu Recht. Wenn sich auch bereits einige allgemeine Hypothesen formulieren lassen, so sind diese Hypothesen doch oftmals unpräzise formuliert und erscheinen in ihrem Gehalt trivial. Das heißt, diese Verallgemeinerungen enthalten kaum etwas, das man nicht schon aufgrund intuitiver, vorwissenschaftlicher Er-

1) Vgl. BESSLER, H.: Aussagenanalyse. Die Messung von Einstellungen im Text der Aussagen der Massenmedien, a.a.O.

2) Dabei kann weitgehend von den bereits vorliegenden Arbeiten des Forschungsprogrammes ausgegangen werden.
Vgl. BESSLER, H.: Zur Theoriebildung in der Massenkommunikationsforschung, a.a.O., und BLEDDIAN, F.: Ergebnisse und neuere theoretische Ansätze über die Wirkungen der Massenkommunikation auf die Einstellungen der Rezipienten, a.a.O.

fahrungen vermutet hätte.¹⁾ Überdies behandeln die meisten der in verschiedenen Hypothesensammlungen enthaltenen Hypothesen nur Einzelaspekte des Massenkommunikationsprozesses, ohne den Versuch zu unternehmen, den gesamten Prozeß in einer umfassenden Theorie in den Griff zu bekommen.²⁾ Ein solcher Versuch ist jedoch nötig, wenn man befriedigende Erklärungen der im Massenkommunikationsprozeß für erklärungsbedürftig gehaltenen Phänomene finden will.

Erste Hinweise für die unter diesem Gesichtspunkt einschlagende Richtung wurden innerhalb des genannten Forschungsprogrammes zunächst in einer "Systematik der Massenkommunikationsforschung" entwickelt.³⁾ In dieser Systematik wird der Prozeß der Massenkommunikation in drei Phasen aufgeteilt: a) der Prozeß der Aussagenentstehung in den Institutionen der Massenkommunikation, b) der Prozeß der Wirkungen von Aussagen der Massenkommunikation auf den einzelnen Rezipienten und c) die Konsequenzen der Beeinflussung einzelner Rezipienten für die Erhaltung und den

1) Trotz gewisser Fortschritte für die Theorienbildung sind auch die bekannten fünf "Verallgemeinerungen" KLAPPERS ein Beispiel hierfür.

Vgl. KLAPPER, J. T.: The Effects of Mass Communication, Glencoe 1960, S. 8. Vgl. hierzu auch BLEDDJIAN, F.: Ergebnisse und neuere theoretische Ansätze über die Wirkungen der Massenkommunikation auf die Einstellungen der Rezipienten, a.a.O., S. 25 - 34

2) Ausführliche Hypothesensammlungen, die sich jedoch fast ausschließlich mit Problemen der Aussagenwirkung beschäftigen, finden sich bei DERELSON, B. und G. A. STEINER: Human Behavior, New York, Chicago und Burlingame 1964, S. 527 - 555 sowie bei COHEN, A.: Attitude Change and Social Influence, New York und London 1964

3) Vgl. BESSLER, H. und F. BLEDDJIAN: Systematik der Massenkommunikationsforschung, München und Basel 1967

Wandel sozialer Systeme. Danach wurde vor allem die empirische¹⁾ und theoretische²⁾ Bearbeitung der ersten und zweiten Phase vorangetrieben, während die dritte Phase, deren Behandlung die Entwicklung eines makrosoziologischen Ansatzes erfordert, erst später untersucht werden soll.

Zur Entwicklung eines theoretischen Konzeptes für die Erklärung der Prozesse der Aussagenentstehung und der Aussagenwirkung wurde, wie bereits gesagt, der Einstellungsbegriff als Basisbegriff ausgewählt. Für diese Auswahl spricht eine Reihe von Gründen, die hier kurz erläutert werden sollen. Der erste Vorteil ist begrifflicher Natur und wird deutlich, wenn man der Charakterisierung des Einstellungsbegriffes von ROTH folgt. ROTH beschreibt Einstellungen mit Hilfe von vier Kriterien und kennzeichnet sie zunächst als hypothetische Konstrukte³⁾, das heißt als nicht unmit-

-
- 1) Eine erste empirische Studie zu Problemen der Aussagenentstehung, bei der ein systemtheoretischer Ansatz verwendet wurde, ist abgeschlossen und wird in Kürze veröffentlicht, vgl. GERBER, C.-P. und M. STOSBERG: Die Massenmedien und die Organisation politischer Interessen, befindet sich im Druck. Ein systemtheoretischer Ansatz liegt auch der Arbeit von ROHL zugrunde; vgl. ROHL, M.: Die Zeitungsredaktion als organisiertes soziales System, Bielefeld 1969.
Die erste empirische Untersuchung aus dem Bereich der Wirkungsforschung steht kurz vor dem Abschluß.
Vgl. STOSBERG, K.: Die Selbsteinschätzung der Rezipienten und die Kommunikationsdiskrepanz als Determinanten des Einstellungswandels im Massenkommunikationsprozeß, a.a.O.
 - 2) Vgl. BESSLER, H.: Zur Theoriebildung in der Massenkommunikationsforschung, a.a.O., sowie BLEDDJIAN, F.: Ergebnisse und neuere theoretische Ansätze über die Wirkungen der Massenkommunikation auf die Einstellungen der Rezipienten, a.a.O.
 - 3) Vgl. ROTH, E.: Einstellung als Determination individuellen Verhaltens, a.a.O., z.B. S. 38.
Der Ausdruck "hypothetisches Konstrukt" ist identisch mit dem in dieser Arbeit verwendeten Terminus "theoretisches Konstrukt". Vgl. auch die Ausführungen hierzu bei BLEDDJIAN, F.: Ergebnisse und neuere theoretische Ansätze über die Wirkungen der Massenkommunikation auf die Einstellungen der Rezipienten, a.a.O., S. 15 - 18

telbar beobachtbare Größen im Sinne der Ausführungen der vorangehenden Abschnitte. Es ist danach möglich, den Konstrukt "Einstellungen" konstitutive Bedeutung durch die Herstellung von Beziehungen mit anderen Konstrukten, die möglicherweise - wie im Falle des hier noch darzustellenden theoretischen Ansatzes - auch Einstellungen sein können, zu verleihen. Des weiteren charakterisiert ROTH Einstellungen durch ihre Objektbezogenheit.¹⁾ Das bedeutet, daß eine verschieden große Anzahl von Personen zu den verschiedensten Objekten unterschiedliche Einstellungen haben kann, die möglicherweise in ihrem Allgemeingrad und in ihrer individuellen Bedeutung variieren können. Durch die Charakterisierung der Einstellungen als objektbezogen wird einerseits hervorgehoben, daß bei einem Versuch, menschliches Verhalten zu erklären, davon auszugehen ist, daß dieses Verhalten immer objektbezogen ist. Zum anderen wird deutlich gemacht, daß für einen solchen Erklärungsversuch nicht die "objektive" Umwelt des Menschen, sondern die subjektiv eventuell unterschiedlich erlebte Umwelt maßgebend sein muß. Das ist, wie BLEDDJIAN hervorhebt, für die Analyse der Wirkungen der Massenkommunikation besonders bedeutsam; die Wirkungen von Aussagen der Massenmedien müssen von dem subjektiven Erleben der Rezipienten ausgehend untersucht werden.²⁾

Als drittes Kriterium zur Beschreibung von Einstellungen nennt ROTH ihre Erworbenheit: Einstellungen werden im Ablauf der individuellen Lerngeschichte erworben und auch

1) Vgl. ROTH, E.: Einstellung als Determination individuellen Verhaltens, a.a.O., S. 62 ff.

2) Vgl.: BLEDDJIAN, F.: Ergebnisse und neuere theoretische Ansätze über die Wirkungen der Massenkommunikation auf die Einstellungen der Rezipienten, a.a.O., S. 17

verändert. Diese Prozesse erfolgen nach den erst teilweise bekannten allgemeinen Lerngesetzen. - Schließlich können Einstellungen anhand ihres Systemcharakters beschrieben werden. Das bedeutet, daß der Änderung einer einzelnen Einstellung im Gesamtgefüge der Einstellungen eines Individuums die Änderung einer oder mehrerer weiterer Einstellungen folgen kann.¹⁾

Anhand dieser vier Merkmale läßt sich eine Reihe von Dimensionen gewinnen, auf denen Einstellungen variieren können.²⁾ Diese Dimensionen sollen hier, den Ausführungen BLEDJIANs³⁾ folgend, kurz aufgezählt werden. Selbst wenn die getrennte Operationalisierung der verschiedenen Dimensionen zur Zeit noch nicht möglich erscheint und bisher noch kaum Versuche in dieser Richtung unternommen wurden, so dürfte doch im Hinblick auf einen Fortschritt in der Theorienbildung und -prüfung auch auf dem Gebiet der Massenkommunikation eine Operationalisierung dieser Dimensionen wünschenswert sein. Der nachfolgend in den für diese Arbeit relevanten Punkten geschilderte theoretische Ansatz orientiert sich jedoch an den derzeitigen meßtechnischen Möglichkeiten und setzt zu seiner Prüfung die Messung von Richtung und Intensität der Einstellungen auf dem Niveau von Ordinalskalen voraus.

-
- 1) Dieser Gedanke läßt sich auch auf den Zusammenhang zwischen den verschiedenen Elementen und Komponenten einer Einstellung ausdehnen. Das soll hier jedoch nicht näher ausgeführt werden.
Vgl. dazu KRECH, D., R. S. CRUTCHFIELD und E. BALLACHEY: *Individual in Society*, New York, San Francisco, Toronto, London und Tokyo 1962, S. 140
 - 2) Vgl. ROTH, E.: *Einstellung als Determination individuellen Verhaltens*, a.a.O., S. 62 ff.
 - 3) Vgl. BLEDJIAN, F.: *Ergebnisse und neuere theoretische Ansätze über die Wirkungen der Massenkommunikation auf die Einstellungen der Rezipienten*, a.a.O., S. 116 - 145

Einstellungen können zunächst in ihrer Richtung variieren; auf dieser Dimension lassen sich positive, neutrale und negative Einstellungen feststellen. Zwei weitere Dimensionen sind Extremität und Intensität. Der Gebrauch dieser Termini in der Literatur ist nicht ganz einheitlich: BLEDJIAN versteht unter Extremität in Anlehnung an HARTLEY und HARTLEY die Entfernung der Position einer Einstellung von einem neutralen Mittelpunkt auf einem Einstellungskontinuum.¹⁾ Unter Intensität wird dann die Stärke verstanden, mit der ein Individuum eine mehr oder weniger extreme Einstellung vertritt. Das nachfolgend mit seinen Determinanten und Resultanten dargestellte theoretische Konzept sieht diese Unterscheidung zwischen Extremität und Intensität von Einstellungen nicht vor; die Prüfung dieses Ansatzes erfordert vielmehr, wie bereits ausgeführt, nur die Messung von Richtung und Intensität von Einstellungen, wobei jedoch unter Intensität im Gegensatz zu der gerade getroffenen Unterscheidung die Entfernung der Position einer Einstellung von einem neutralen Mittelwert zu verstehen ist. Die weiteren Dimensionen, auf denen Einstellungen variieren können, werden bei diesem Ansatz aus vorwiegend meßtechnischen Gründen nicht berücksichtigt.

Einstellungen können nach dem Grad ihrer Bedeutsamkeit bzw. Zentralität unterschieden werden: Mit Zentralität ist die jeweilige Position einer Einstellung zu einem bestimmten Einstellungsobjekt innerhalb der gesamten Einstellungshierarchie eines Individuums gemeint. Man spricht, um zwischen Bedeutsamkeit und Zentralität zu differenzieren,

1) Vgl. BLEDJIAN, F.: Ergebnisse und neuere theoretische Ansätze über die Wirkungen der Massenkommunikation auf die Einstellungen der Rezipienten, a.a.O., S. 118
Vgl. HARTLEY, E. L. und R. E. HARTLEY: Die Grundlagen der Sozialpsychologie, Berlin 1955, S. 445 ff.

davon, daß eine zentrale Einstellung dauerhaft ist, während eine Einstellung geringer Zentralität durch Aktualisierung bedeutsam werden kann.¹⁾ Weitere Dimensionen, auf denen Einstellungen variieren können, sind die Komplexität, die Differenziertheit und die Verbundenheit von Einstellungssystemen.²⁾ Die Komplexität eines Einstellungsgefüges wird durch die Anzahl der Einstellungen in diesem Gefüge charakterisiert. Unter Differenziertheit ist der Grad der Trennung zwischen einzelnen Einstellungen oder Einstellungsbestandteilen zu verstehen. Hohe Differenziertheit beschreibt BLEDDJIAN am Beispiel eines Rezipienten, der bei Studentendemonstrationen klar zwischen Demonstranten der verschiedenen Hochschulgruppen, zwischen verfassungskonformen und verfassungswidrigen Handlungen etc. unterscheidet. Die dritte Dimension schließlich bezeichnet das Ausmaß der Verbundenheit zwischen den Einstellungen in einem Einstellungsgefüge. Beispiele für eine hohe Verbundenheit von Einstellungen in einem System finden sich, wie BLEDDJIAN ausführt, in der Untersuchung von ADORNO u.a.³⁾ Über die autoritäre Persönlichkeit: Hier zeigte sich, daß bestimmte Einstellungen gegenüber Juden, Negern und anderen Minoritäten sowie gegenüber den Vereinigten Staaten sogenannte "attitude clusters" bilden und zudem noch mit einer faschistischen Einstellung korrelieren.⁴⁾ Schließlich erläutert BLEDDJIAN die Dimension der

-
- 1) Vgl. BLEDDJIAN, F.: Ergebnisse und neuere theoretische Ansätze über die Wirkungen der Massenkommunikation auf die Einstellungen der Rezipienten, a.a.O., S. 123 f.; siehe dort auch weitere Literatur
 - 2) Vgl. ders., a.a.O., S. 132 ff.
Der dieser Unterscheidung zugrunde liegende Gedanke kann auch auf die Komponenten bzw. Elemente von Einstellungen übertragen werden.
 - 3) Vgl. ADORNO, T. W., E. FRENKEL-BRUNSWIK, D. J. LEVINSON und R. M. SANFORD: The Authoritarian Personality, New York 1950
 - 4) Vgl. aber hierzu die Kritik HYMANs an der Vorgehensweise der Studie über die autoritäre Persönlichkeit; HYMAN, H.: Survey Design and Analysis, New York 1955, S. 181 f.

K o n s i s t e n z , die sich auf die subjektive Vereinbarkeit zwischen den Bestandteilen eines Einstellungssystems bezieht. - Alle diese zuletzt genannten Dimensionen werden in dem kurz darzustellenden theoretischen Konzept - wie gesagt - noch nicht berücksichtigt. Dieses Konzept bietet im wesentlichen ein System von Determinanten und Resultanten für die Formulierung von Hypothesen über den Prozeß der Aussagenentstehung und der Aussagenwirkung. Gerade mit der Dimension der Konsistenz und den auf dem Konsistenzprinzip aufbauenden Theorien dürfte jedoch ein Ansatz zur Ausfüllung des Konzeptes und zur Formulierung konkreter Hypothesen gegeben sein.¹⁾

Bisher wurden hauptsächlich die begrifflichen Vorteile des Einstellungskonstruktes bei der Verwendung in Hypothesen, die menschliches Verhalten erklären sollen, aufgezählt. Damit wurden die Möglichkeiten erörtert, dem Konstrukt der Einstellungen konstitutive Bedeutung zu verleihen. Für die Verwendung des Einstellungsbegriffes sprechen jedoch noch eine Reihe weiterer Umstände: Der Einstellungsbegriff dürfte im Vergleich zu anderen konkurrierenden Konstrukten trotz der auftretenden Schwierigkeiten relativ leicht zu operationalisieren sein. (Mit den Techniken der Messung von Einstellungen, die sowohl bei Kommunikatoren als auch bei Rezipienten angewendet werden können, beschäftigt sich das Kapitel 4 dieser Arbeit.) Ein eher praktischer Grund für die Verwendung des Einstellungsbegriffes liegt in dem Umstand, daß viele der bereits vorliegenden und geprüften Hypothesen der Massenkommunikationsforschung mit dem Einstellungsbegriff oder verwandten Konstrukten wie z.B. "opinion" oder Prädispositionen arbeiten.²⁾ Wenn nun ein umfassender

1) Vgl. die ausführliche Diskussion der Konsistenztheorien bei BLEDDIAN, F.: Ergebnisse und neuere theoretische Ansätze über die Wirkungen der Massenkommunikation auf die Einstellungen der Rezipienten, a.a.O., S. 146 ff.

2) Vgl. BESSLER, H.: Zur Theoriebildung in der Massenkommunikationsforschung, a.a.O., S. 29 f. BESSLER bezieht sich dabei auf die bereits zitierte Hypothesensammlung von BERELSON, B. und G. A. STEINER: Human Behavior, a.a.O., S. 527 ff.

theoretischer Ansatz den Einstellungsbegriff als Basisbegriff wählt, so ist es möglich, viele Einzelhypothesen, deren Variablen Einstellungen oder ähnliche Konstrukte sind, auf ihre Konsistenz mit dem umfassenden Ansatz zu prüfen. Außerdem können von den Einzelhypothesen möglicherweise zahlreiche Anregungen für die Formulierung von Hypothesen im Rahmen eines allgemeinen Ansatzes zur Erklärung von Massenkommunikationsprozessen ausgehen.

Schließlich scheint der Einstellungsbegriff auch in der allgemeinen sozialpsychologischen Diskussion zunehmend an Bedeutung zu gewinnen; das zeigt sich nicht zuletzt in den bereits erwähnten konsistenztheoretischen Ansätzen. Der Vorteil der Verwendung des Einstellungsbegriffes in einer Theorie der Massenkommunikation besteht dann darin, daß diese Gemeinsamkeit eine Verknüpfung massenkommunikationstheoretischer Ansätze mit allgemeinen Theorien menschlichen Verhaltens erleichtert.¹⁾

Was den im folgenden Abschnitt erläuterten theoretischen Ansatz betrifft, so hat die Verwendung des Einstellungsbegriffes noch einen weiteren ganz speziellen Vorteil: Alle Variablen werden als Einstellungen formuliert; dadurch, daß die Resultante der Hypothesen über die Aussagenentstehung gleichzeitig eine der Determinanten der Hypothesen über die Aussagenwirkung ist, besteht die Möglichkeit einer Verbindung der Hypothesen über die Aussagenentstehung und die Aussagenwirkung.²⁾ Eine solche Verbindung wäre nicht möglich oder zumindest erschwert, wenn in den Hypothesen nicht einheitlich Einstellungen als Variablen verwendet würden.

1) Diese Frage soll hier nicht weiter erörtert werden; vgl. BESSLER, H.: Zur Theoriebildung in der Massenkommunikationsforschung, a.a.O., S. 33 ff.

2) Vgl. ders., a.a.O., S. 39 und 83

242 Einstellungen als Determinanten und Resultanten eines theoretischen Konzeptes zur Erklärung von Massenkommunikationsprozessen

Die Darstellung des theoretischen Ansatzes, der im Rahmen des bereits mehrfach erwähnten Forschungsprogrammes zur Massenkommunikation entwickelt wurde, soll in der hier vorliegenden Arbeit in aller Kürze erfolgen.¹⁾ Es ist lediglich aufzuzeigen, welche Variablen dieses Ansatzes mit Hilfe der nachfolgend in Kapitel 4 diskutierten Techniken zu messen sind. In diesem Zusammenhang soll daher keine nähere Begründung für die Auswahl der in dem Ansatz verwendeten Variablen aus anderen, konkurrierenden Variablen gegeben werden; es wäre damit nur eine Wiederholung von bereits an anderer Stelle ausgeführten Argumenten verbunden.²⁾ Es sei nur darauf hingewiesen, daß die Auswahl und Formulierung der Variablen aufgrund eines intensiven Studiums der bisherigen Forschungsergebnisse unter meßtechnischen Gesichtspunkten erfolgte; dabei hat nicht zuletzt auch die Frage nach den möglichen Zeitpunkten für Messungen im Ablauf des Massenkommunikationsprozesses eine Rolle gespielt. Ausdrücklich zu betonen ist ferner die Vorläufigkeit des Konzeptes: In empirischen Untersuchungen muß es auf seine Brauchbarkeit geprüft werden.

1) Die ausführliche Darstellung findet sich bei BESSLER, H.: Zur Theoriebildung in der Massenkommunikationsforschung, a.a.O., S. 72 - 97

2) Vgl. ebenda, den Abschnitt "Potentielle Determinanten und Resultanten von Hypothesen über Wirkungen der Massenkommunikation", S. 50 ff. und in bezug auf die Aussagenentstehung S. 82 ff.

Ein auf dem Einstellungsbegriff basierender theoretischer Ansatz kann nicht alle für erklärungsbedürftig gehaltenen Verhaltensweisen von Kommunikatoren und Rezipienten erklären: "Das Zustandekommen von Aussagen, in denen Einstellungen zu bestimmten Einstellungsobjekten vertreten werden, und das Zustandekommen von Einstellungsänderungen bei den Rezipienten dieser Aussagen sind jedoch erklärungsbedürftige Phänomene von vorrangiger Bedeutung in der Massenkommunikationsforschung."¹⁾ Für die Bildung von Hypothesen zur Erklärung des Prozesses der A u s s a g e n e n t - s t e h u n g sieht das Konzept folgende Variablen vor:

- I. " D e t e r m i n a n t e n des Prozesses der Aussagenentstehung sind die Einstellungen der produzierenden Kommunikatoren, die in der Vorproduktionsphase mit Einstellungsskalen zu messen sind, und zwar:
 1. (X_A), ihre Einstellungen zu den Einstellungsobjekten 1, 2, 3 ... n, die sie zum Gegenstand ihrer Produktionsaufgabe zählen;
 2. (X_B), ihre Erwartungen von der Durchsetzbarkeit ihrer Einstellungen zu den Einstellungsobjekten 1, 2, 3 ... n, die sie zum Gegenstand ihrer Produktionsaufgabe zählen.
- II. R e s u l t a n t e n des Prozesses der Aussagenentstehung sind (Y), die in den veröffentlichten Aussagentexten²⁾ tatsächlich geäußerten Einstellungen zu den

1) BESSLER, H.: Zur Theoriebildung in der Massenkommunikationsforschung, a.a.O., S. 39

2) In dieser Formulierung ist nur von "Aussagentexten" und nicht allgemein von "Aussagen" die Rede, weil - wie BESSLER aufzeigt - beim heutigen Stand der Forschung die Messung von Einstellungsäußerungen mit aussagenanalytischen Mitteln in befriedigender Weise nur für vorliegende Texte möglich ist.
Vgl. BESSLER, H.: Zur Theoriebildung in der Massenkommunikationsforschung, a.a.O., S. 8 f. und S. 58.
Das gleiche gilt für die Formulierung der Variablen zur Bildung von Hypothesen über die Aussagenwirkungen.

Einstellungsobjekten I, II, III ... N, die nach der Veröffentlichung der Aussagen mit aussagenanalytischen Maßnahmen zu messen sind."¹⁾

Für den Prozeß der Aussagenentstehung gilt demnach die Funktion $Y = F [(X_A), (X_B)]$.

Zur Erklärung des Prozesses der Aussagenwirkung werden folgende Variablen in den Hypothesen des Konzeptes verwendet:

I. " D e t e r m i n a n t e n des Wirkungsprozesses sind

1. (X_A) , die in den Aussagentexten tatsächlich geäußerten Einstellungen zu den Einstellungsobjekten I, II, III ... N, die in der präkommunikativen Phase mit aussagenanalytischen Techniken zu messen sind;
2. die Einstellungen der Rezipienten, die in der präkommunikativen Phase mit Einstellungsskalen zu messen sind, und zwar
 - a) (X_B) , ihre Einstellungen zu den Einstellungsobjekten I, II, III ... N, zu denen in den Aussagentexten tatsächlich Stellung genommen wird;
 - b) (X_C) , ihre Selbsteinschätzung²⁾ hinsichtlich ihrer Einstellungen zu den Einstellungsobjekten I, II, III ... N, zu denen in den Aussagentexten tatsächlich Stellung genommen wird;
 - c) (X_D) , ihre Erwartungen von den mutmaßlichen Einstellungsäußerungen in den Aussagentexten, sofern diese die Einstellungsobjekte I, II, III ... N

1) BESSLER, H.: Zur Theoriebildung in der Massenkommunikationsforschung, a.a.O., S. 94

2) Für die Brauchbarkeit dieser Variable sprechen die Ergebnisse der Untersuchung von STOSBERG, K.: Die Selbsteinschätzung der Rezipienten und die Kommunikationsdiskrepanz als Determinanten des Einstellungswandels im Massenkommunikationsprozeß, a.a.O.

betreffen, zu denen in den Aussagen tatsächlich Stellung genommen wird.

II. R e s u l t a n t e n des Wirkungsprozesses sind $(Y_A) \dots (Y_N)$, die Einstellungsänderungen, die sich aus einem Vergleich der mit Einstellungsskalen in der präkommunikativen Phase gemessenen Einstellungen mit den in der postkommunikativen Phase gemessenen Einstellungen der Rezipienten zu den Einstellungsobjekten I, II, III ... N ergeben, zu denen in den Aussagentexten tatsächlich Stellung genommen wird.¹⁾

Die Hypothesen zur Erklärung der Aussagenwirkungen lassen sich nach diesem Konzept in die Form der Funktion $(Y_A) \dots (Y_N) = F(X_A), (X_B), (X_C), (X_D)$ bringen. Eine Verknüpfung der beiden Ansätze zur Erklärung von Aussagenentstehung und Aussagenwirkung dürfte, wie gesagt, relativ leicht aufgrund der Tatsache zu erreichen sein, daß die Resultante (Y) der Hypothesen über die Aussagenentstehung identisch ist mit der Determinante (X_A) der Hypothesen über die Aussagenwirkung.

Von den in diesen beiden Ansätzen verwendeten acht Variablen sind sechs mit Hilfe von Einstellungsskalen zu erfassen: in den Hypothesen über die Aussagenentstehung die Variablen (X_A) und (X_B) und in den Hypothesen über die Aussagenwirkung die Variablen (X_B) , (X_C) , (X_D) und $(Y_A) \dots (Y_N)$. Wenn konkrete Hypothesen über die Beziehungen zwischen

1) BESSLER, H.: Zur Theoriebildung in der Massenkommunikationsforschung, a.a.O., S. 75 f.
Bei diesen beiden Ansätzen handelt es sich streng genommen noch nicht um ein theoretisches Konzept, sondern eigentlich erst um ein Rahmenprogramm für die Hypothesenformulierung mit Hilfe der genannten Variablen. Eine Ausfüllung könnte, wie bereits ausgeführt, etwa anhand der Konsistenztheorien erfolgen. Ob die in der Arbeit von BESSLER vorgeschlagene Forschungsstrategie erfolgversprechend im Hinblick auf den Erkenntnisfortschritt in der Massenkommunikationsforschung ist, erscheint allerdings aus Gründen, die hier nicht näher erörtert werden können, zumindest zweifelhaft.

allen acht Variablen formuliert sind, so hat deren Prüfung mit Hilfe einer koordinierten Messung der Variablen zu erfolgen.¹⁾ Das bedeutet, daß sowohl die genannten Einstellungen der Kommunikatoren und Rezipienten als auch die in den Aussagentexten geäußerten Einstellungen mittels Einstellungsskalen bzw. aussagenanalytischen Techniken hinsichtlich Richtung und Intensität auf dem Niveau von Ordinalskalen gemessen werden müssen. BESSLER hat in seinen Arbeiten nachgewiesen, daß es Verfahren der Aussagenanalyse gibt, mit denen eine solche Messung für Einstellungen, die in Texten geäußert werden, möglich ist. Im Kapitel 4 der vorliegenden Arbeit wird eine Reihe von Skalierungsverfahren erläutert, welche die Messung von Richtung und Intensität der Einstellungen von Kommunikatoren und Rezipienten auf Ordinalskalenniveau erlauben. Zuvor soll jedoch noch auf die grundsätzlichen Probleme der Messung eingegangen werden, wie sie für die Operationalisierung der Determinanten und Resultanten des hier zur Diskussion stehenden Ansatzes von unmittelbarer Bedeutung sind.

1) Vgl. BESSLER, H.: Zur Theoriebildung in der Massenkommunikationsforschung, a.a.O., S. 78.

3 Grundlagen der Messung

31 Definitionsversuche: Messung und Skalierung

Wissenschaftliche Beobachtungen beziehen sich auf Eigenschaften, die irgendwelche Objekte besitzen. Als Eigenschaften sollen hier die mehr oder weniger gut beobachtbaren Charakteristika von Dingen oder Ereignissen bezeichnet werden; erst durch ihre Eigenschaften werden die Objekte erfahrbare. Die Unterscheidung zwischen Eigenschaften und Objekten ist insofern von Bedeutung, als immer nur bestimmte Eigenschaften in der Realität gemessen werden können, wenn man etwas über Objekte aussagen will.¹⁾ In diesem Zusammenhang spricht SCHEUCH davon, "daß Messen eine Abstraktion von der Konkretheit des Objekts voraussetzt."²⁾

In diesem Sinne erscheint der oft vertretene Einwand gegen die Möglichkeiten quantifizierender Sozialforschung, der Mensch sei nicht meßbar, unsinnig. Nicht der Mensch als Objekt soll gemessen werden, sondern immer nur bestimmte, theoretisch relevante Eigenschaften des Menschen. Auch in den Naturwissenschaften werden nicht Körper gemessen, sondern lediglich die Länge, das Gewicht, die Temperatur derselben. Welche Eigenschaften jeweils gemessen werden sollen, kann nur mit Hilfe einer entsprechenden Theorie entschieden werden. Das gilt analog auch für die vorliegende Arbeit: Sie beschäftigt sich mit der Messung von Einstellungen als einer theoretisch relevanten Variablen, die u.a. bei der Formulierung von Hypothesen zur Er-

1) TORGERSON macht in diesem Zusammenhang die Unterscheidung zwischen "property" und "system". Die Verwendung des Terminus "system" erscheint jedoch hier etwas unglücklich, weil gerade der Begriff des Systems auch in verschiedenen anderen Bedeutungen benutzt wird. Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 9

2) SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, in: R. KÜNIG (Hrsg.): Handbuch der empirischen Sozialforschung, Band 1, 2. Aufl., Stuttgart 1967, S. 348

klärung der Entstehung und der Wirkung von Aussagen der Massenkommunikation verwendet werden können.

Im Frühstadium einer wissenschaftlichen Disziplin sind viele Eigenschaftsbegriffe klassifikatorisch. Mit fortschreitender Entwicklung werden jedoch die meisten dieser Begriffe durch Ordnungsbegriffe oder durch metrische Begriffe ersetzt.¹⁾ Damit ist eine Reihe von Vorteilen verbunden, bei deren Aufzählung wiederum der enge Zusammenhang zwischen Messung und Theoriebildung zum Ausdruck kommt:²⁾

- a) Durch die Einführung von Ordnungsbegriffen oder metrischen Begriffen wird es möglich, zwischen den Fällen einer bestehenden Klasse zu differenzieren und damit eine größere Flexibilität und Feinheit bei der Beschreibung zu erreichen.
- b) Die Benutzung solcher Begriffe verdeutlicht außerdem die relative Position verschiedener Aussagen in der durch den betreffenden Begriff gebotenen Rangfolge. HEMPEL belegt das an einem Beispiel: Bei einer metrischen Begriffsbildung ist das Verhältnis von zwei Windstärken - etwa 30 Meilen und 18 Meilen pro Stunde - genau festgelegt. Mit Hilfe der Klassifikation "leichte Brise" und "mäßige Brise" kann eine solche Aussage jedoch nicht ohne weiteres gemacht werden.³⁾
- c) Größere Flexibilität bei der Beschreibung (vgl. a) bedeutet auch größere Flexibilität bei der Formulierung allgemeiner Gesetze.

1) Vgl. hierzu auch den nachfolgenden Abschnitt 32 über das Niveau der Messung.

2) Vgl. HEMPEL, C. G.: Fundamentals of Concept Formation in Empirical Science, a.a.O., S. 56 f.

3) Vgl. HEMPEL, C. G.: Fundamentals of Concept Formation in Empirical Science, a.a.O., S. 57

- d) Die Einführung metrischer Begriffe ermöglicht die Anwendung von Begriffen und Modellen der höheren Mathematik.

Was bedeutet nun eigentlich Messung und Skalierung? Gemäß der bisher verwendeten Terminologie kann unter Messung die Konstruktion einer bestimmten Art von Korrespondenzregeln verstanden werden. Diese Korrespondenzregeln sehen vor "die Zuordnung von Zahlen (numbers), um Eigenschaften zu repräsentieren"¹⁾. Alle späteren, zum Teil ausführlicheren und präziseren Definitionsversuche knüpfen in etwa an dieser Definition CAMPBELLS an. Ohne auf die Entwicklung dieses Begriffes im einzelnen einzugehen²⁾, seien hier lediglich einige neuere Definitionsversuche aufgezeigt. Die in der Literatur meistzitierte Definition stammt von STEVENS: "Im weitesten Sinne ist Messung die Zuordnung von Zahlzeichen (numerals) zu Objekten oder Ereignissen nach Regeln"³⁾. SCHEUCH definiert folgendermaßen: "Unter Messen sei verstanden die nach einer expliziten Regel erfolgende Zuordnung von Zahlenwerten zur Wiedergabe solcher Eigenschaften von Objekten, die nicht selbst numerisch sind"⁴⁾. In den bisher angeführten Definitionen werden die Ausdrücke "number" und "numeral" verwendet. "Numerals" sind dabei lediglich

-
- 1) CAMPBELL, N. R.: What is Science, New York 1921, S. 110
- 2) Vgl. WOOLF, H. (Hrsg.): Quantification. A History of the Meaning of Measurement in the Natural and the Social Sciences, Indianapolis und New York 1961. Zur Geschichte der Messung im Bereich der Soziologie vgl. in diesem Buch den Artikel von LAZARSFELD, P. F.: Notes on the History of Quantification in Sociology - Trends, Sources and Problems, S. 147 - 203.
- 3) STEVENS, S. S.: Mathematics, Measurement and Psychophysics, a.a.O., S. 1
- 4) SCHEUCH, E. K.: Skalierungsverfahren als Instrument der Sozialforschung, Meisenheim a. G. 1967, noch nicht erschienen; so zitiert bei: SCHEUCH, E. K.: Skalierung, in: R. KÖNIG (Hrsg.): Handbuch der empirischen Sozialforschung, a.a.O., im Anhang, S. 738

Symbole der Form 1, 2, 3 ..., ohne daß mit diesen Symbolen zunächst eine quantitative Bedeutung verbunden wäre; "numeral" bedeutet also Zahlzeichen. Wird ein solches Zahlzeichen mit quantitativer Bedeutung ausgestattet, so spricht man von "number".¹⁾

Wenn gemäß der Definition STEVENS Messung die Zuordnung von Zahlzeichen zu Objekten oder Ereignissen ist, so ist es entscheidend für die Qualität einer Messung, nach welchen Regeln die Zuordnung erfolgt. Bei einer Messung wird die Absicht verfolgt, "daß wir Objektrelationen, die nicht unsere Erfindung sind, durch Zahlenrelationen abbilden, die unsere Erfindung sind"²⁾. Das Ausmaß des "Meßfehlers" entscheidet darüber, "ob es sinnvoll ist, die zugeordneten Zahlen als Meßwerte der Objekte zu bezeichnen"³⁾. Diese Überlegung kann anhand der folgenden Gleichung verdeutlicht werden⁴⁾:

$$V_t = V_w + V_e$$

V_t = Varianz der ausgeteilten Zahlen

V_w = Varianz zu Lasten der wirklichen Unterschiede zwischen den Objekten

V_e = Varianz zu Lasten des Meßfehlers

Im Sinne dieser Gleichung ist es also nur dann sinnvoll, zugeordnete Zahlen als Abbild von Objektrelationen zu betrachten, wenn V_e im Verhältnis zu V_w einen bestimmten, innerhalb des jeweils verwendeten Meßmodells festzulegenden Wert, nicht überschreitet. Eine Abbildung von

1) Vgl. KERLINGER, F. N.: Foundations of Behavioral Research, New York, Chicago, San Francisco, Toronto und London 1964, S. 413

2) SIXTL, F.: Meßmethoden in der Psychologie, Weinheim 1967, S. 3

3) ders., a.a.O., S. 2

4) ebenda

Objektrelationen durch Zahlenrelationen ist dann möglich, "wenn bestimmte Eigenschaften der Zahlen *i s o m o r p h* zu bestimmten Eigenschaften der Objekte sind"¹⁾. In Anlehnung u.a. an SUPPES und ZINNES²⁾ gibt SIXTL folgende Definition von Isomorphie: "Gibt es eine eindeutige Abbildung f von S auf S' derart, daß für $x \leq y$ auch $f(x) \leq f(y)$ und für $f(x) \leq f(y)$ auch $x \leq y$, so ist f eine isomorphe Abbildung, beziehungsweise S und S' heißen isomorph. Es sei nun S' ein numerisches Relationssystem B , das einem empirischen A isomorph ist: B heißt dann ein Modell von A ".³⁾ Ob ein Modell für einen bestimmten Objektbereich angemessen ist, kann nur durch empirische Prüfungen entschieden werden; logische Gewißheit ist nicht zu erlangen.⁴⁾ Damit ist wiederum ein Aspekt des Validitätsproblems angesprochen, das bereits als das Kernproblem jeder Messung vor allem beim gegenwärtigen Stand der Sozialwissenschaften charakterisiert wurde.

Bezogen sich die bisherigen Ausführungen auf Messung allgemein, so kann nunmehr der Prozeß der Skalierung als eine bestimmte Form des Messens bezeichnet werden, als diejenige Form nämlich, "die durch standardisierte Kombination von Beobachtungen zu einer geometrischen Repräsentation von intendierten bzw. latenten Klassifikationen führt".⁵⁾ SCHEUCH stellt insbesondere folgende Anforderungen an

- 1) SIXTL, F.: Meßmethoden in der Psychologie, a.a.O., S. 3
- 2) Vgl. SUPPES, P. und J. L. ZINNES: Basic Measurement Theory, in: R. J. LUCE, R. R. BUSI, und E. GALANTER (Hrsg.): Handbook of Mathematical Psychology, New York und London 1963, S. 4 - 6
- 3) SIXTL, F.: Meßmethoden in der Psychologie, a.a.O., S. 3
- 4) Mit der Konstruktion solcher Modelle und der Prüfung ihrer Angemessenheit beschäftigt sich das Kapitel 4 der vorliegenden Arbeit.
- 5) SCHEUCH, E. K.: Skalierung, a.a.O., S. 73ff; SCHEUCH entwickelt diese Definition in Anlehnung an COOMBS, C. H.: A Theory of Data, New York, London und Sidney, 1964

einen Meßvorgang, damit man von Skalierung sprechen kann:
"1. Es werden Objekte nach den an ihnen beobachteten Eigenschaften auf Grund einer interpersonell konstanten Vorgehensweise auf einem Kontinuum plaziert, wobei 2. ein Modell der Verteilungsgesetze für die Indikatoren, ihre Interaktion und die Verteilung der Objekte vorausgesetzt wird, 3. objektive Kriterien zur Beurteilung des Grades an Übereinstimmung zwischen Modell-Annahmen und Beobachtungsmaterial vorliegen, sowie 4. auf die Verwendung eines äußeren Maßstabes verzichtet werden kann und die Ordnung lediglich aus dem Beobachtungsmaterial heraus erfolgt."¹⁾ Die Bedeutung der einzelnen Anforderungen wird bei der Diskussion der verschiedenen Skalierungstechniken in Kapitel 4 deutlich. Um sich über die Angemessenheit eines Modells klar werden zu können, müssen zunächst die mit den Fragen des Meßniveaus und der Art der Messung zusammenhängenden Probleme behandelt werden.

32 Das Niveau der Messung

Eine Reihe realer Zahlen kann mit Hilfe der folgenden drei Kriterien beschrieben werden²⁾:

- a) Die Zahlen einer Reihe können geordnet sein (Ordnung);
- b) die Abstände zwischen den Zahlen einer Reihe können geordnet sein (Abstand);
- c) die Reihe kann einen Ursprung besitzen, der mit der Zahl "null" bezeichnet wird (Ursprung).

Der Einfachheit halber wird im folgenden ebenso wie bei TORGERSON nur von den Kriterien der Ordnung, des Abstandes und des Ursprungs gesprochen.

1) SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, a.a.O., S. 352

2) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 15

Skalen können nun danach klassifiziert werden, wieweit den genannten Kriterien im Hinblick auf die von ihnen zu messenden Eigenschaften von Objekten empirische Bedeutung zukommt oder nicht. Bei einer vorliegenden Skala können ein, zwei, drei oder auch keines dieser Kriterien empirische Bedeutung erhalten. Auf diese Weise kommt man zu einer Unterscheidung der verschiedenen Meßniveaus.¹⁾ Eine Entscheidung darüber, welchen der Kriterien empirische Bedeutung zukommt, ist einerseits eine Frage der Definition, andererseits eine Frage der empirischen Prüfung, ob die Annahmen (Postulate, Axiome), die bei einem bestimmten Meßniveau zugrunde gelegt werden, auch tatsächlich vorliegen. Die Frage, auf welche Art die Kriterien der Ordnung, des Abstandes und des Ursprunges Bedeutung gewinnen, führt zu einer weiteren Unterscheidung nach verschiedenen Arten der Messung.²⁾

Die in diesem Abschnitt zu behandelnden Skalentypen bilden von der Nominal- bis zur Ratioskala ein zunehmend höheres Niveau des Messens,³⁾ das durch Inzufügen jeweils neuer Axiome erreicht wird. Die Klassifikation beruht darauf, wieviel Information über die zu messende Eigenschaft jeweils gegeben wird. In den meisten Lehrbüchern, die sich mit dieser Frage beschäftigen, werden herkömmlicherweise in Anlehnung an STEVENS⁴⁾ vier verschiedene Meßniveaus unterschieden: Nominal-, Ordinal-, Intervall- und Ratioskala. Mit Hilfe der genannten Kriterien lassen sich jedoch verschiedene Zwischentypen entwickeln, die in diesem Abschnitt

-
- 1) TORGERSON spricht von "types of scales".
Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 15
 - 2) Vgl. den nachfolgenden Abschnitt 33
 - 3) Eine gewisse Einschränkung erfährt diese Behauptung bei der Erörterung der geordneten metrischen Skala; vgl. unten Punkt e.
 - 4) Vgl. STEVENS, S. S.: Mathematics, Measurement, and Psychophysics, a.a.O., S. 23 ff.

mitbehandelt werden, da sie eventuell gerade in den Sozialwissenschaften bei der Messung von Einstellungen mit den dort zur Verfügung stehenden Instrumenten besondere Bedeutung erlangen können.¹⁾ Eine Klärung der Frage, welches Meßniveau mit einem bestimmten Instrument erreicht werden kann, ist besonders darum wichtig, weil eine Entscheidung über die möglichen Rechenoperationen und Transformationen nur auf der Grundlage eines bestimmten Meßniveaus zu fällen ist.

a) Die Nominalskala

Für die Nominalskala als niedrigstem Meßniveau trifft keines der genannten Kriterien der Ordnung, des Abstandes und des Ursprunges zu. Es handelt sich hier lediglich um einfaches Kategorisieren, also um eine Einteilung einer Menge von Objekten in einzelne, eindeutig voneinander unterscheidbare Klassen im Hinblick auf eine bestimmte Eigenschaft: Die Objekte werden danach klassifiziert, ob sie eine bestimmte (die zu "messende") Eigenschaft aufweisen oder nicht. Es könnten demnach ebenso gut anstelle von Zahlen andere Symbole, z.B. Namen, zugeordnet werden.

Der Nominalskala liegen die Identitätspostulate zugrunde:

1. Entweder $i = j$ oder $i \neq j$. Das bedeutet, daß die Objekte und die ihnen zugeordneten Zahlen entweder identisch sind oder nicht.
2. Wenn $i = j$ dann $j = i$ (Symmetriepostulat).
3. Wenn $i = j$ und $j = k$ dann $i = k$ (Transitivitätspostulat).

Als Rechenoperationen kommen auf dem Niveau der Nominalskalen in Frage: Zählen, die Bildung von Häufigkeiten sowie die Berechnung von Modus und Kontingenzkoeffizient.

1) Eine gegensätzliche Ansicht vertritt SIXTL, F.: Meßmethoden in der Psychologie, a.a.O., S. 9

Als Transformationen sind alle systematischen Operationen möglich, soweit die genannten Postulate nicht verletzt werden.

Im Verlauf der Diskussion der verschiedenen Meßniveaus hat es eine Reihe von Meinungsverschiedenheiten darüber gegeben, ob man bei einer Nominalskala bereits von "Messen" sprechen könne. So vertritt TORGERSON¹⁾ die Ansicht, daß bei einer Nominalskala noch kein Messen vorliege, weil sich bei einer Messung die Zuordnung der Zahlen auf die Ausprägungen einer Eigenschaft von Objekten beziehen müsse; im Falle der Nominalskala dagegen beziehe sich diese Zuordnung auf die Objekte direkt. Bei einer Nominalskala würden Objekte klassifiziert, Messung beziehe sich jedoch auf das Ausmaß, in dem eine Eigenschaft vorhanden sei. Demgegenüber ist einzuwenden, daß auch die im Falle der Nominalskala erfolgende Klassifikation von Objekten keinen direkten Bezug auf die Objekte selbst nimmt; eine solche Klassifikation erfolgt vielmehr danach, ob bestimmte Eigenschaften vorhanden sind oder nicht. Es spricht folglich nichts dagegen, im Widerspruch zu der Auffassung TORGERSONs auch beim Vorliegen einer Nominalskala von Messung im hier definierten Sinne zu sprechen. Die von TORGERSON verwendeten Kriterien der Ordnung, des Abstandes und des Ursprunges schließen es nicht aus, auch einen Skalentyp zu bilden, bei dem diese Kriterien eben nicht vorliegen: eben den Typ der Nominalskala.

1) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 17

Die damit vertretene Auffassung kann zur Klärung eines in der wissenschaftstheoretischen Diskussion immer noch umstrittenen Problems beitragen: Vielfach wird behauptet, daß es in den Sozialwissenschaften Phänomene gebe, die sich einer quantitativen Erfassung entzögen und lediglich qualitativer Betrachtung zugänglich seien. Die mit einer solchen Behauptung verbundene essentielle Unterscheidung soll zumeist einen methodologischen Autonomieanspruch der Sozialwissenschaften begründen. Nach der in dieser Arbeit vertretenen Ansicht handelt es sich jedoch hier nicht um einen essentiellen Unterschied, wie bereits an anderer Stelle betont wurde.¹⁾ Bei der Unterscheidung zwischen Qualität und Quantität dürfte es sich vielmehr um einen Unterschied im Meßniveau handeln; denn bei einer sogenannten qualitätsmäßigen Betrachtung liegt nichts anderes als eine klassifikatorische Begriffsbildung vor, also das Meßniveau der Nominalskala, während bei einer sogenannten quantitätsmäßigen Betrachtung höhere Meßniveaus, wie sie im folgenden zu erläutern sind, benutzt werden. Die Vorteile der Einführung von Ordnungsbegriffen und von metrischen Begriffen auch in den Sozialwissenschaften, also die Benutzung höheren Meßniveaus, wurden bereits in Abschnitt 31 aufgezeigt.

Qualitätsmäßige Betrachtung heißt demnach Verwendung des niedrigsten Meßniveaus; nichts deutet darauf hin, daß sich auf diese Weise erfaßte Objekte grundsätzlich der Betrachtung bestimmter ihrer Eigenschaften auf einem anderen Meßniveau entzögen. Ein essentieller Unterschied zwischen Qualität und Quantität kann aus diesem Grunde ebenso wenig anerkannt werden wie ein methodologischer Autonomieanspruch

1) Vgl. oben Abschnitt 23

der Sozialwissenschaften. Es ist lediglich möglich, daß bestimmte in den Sozialwissenschaften relevante Variablen im Augenblick noch nur auf Nominalskalenniveau gemessen werden können. Aber: "Wenn sich beim gegenwärtigen Stand der Sozialforschung irgendwelche theoretisch relevante Variablen einer sinnvollen Zuordnung von quantitativen Größen zur Kennzeichnung ihrer Eigenschaften entziehen, so liegt das nicht an einer Immunität dieser Qualitäten in bezug auf eine quantitative Betrachtung, sondern an den Unzulänglichkeiten der bisher entwickelten sozialwissenschaftlichen Forschungstechniken."¹⁾

b) Die teilweise geordnete Skala

Die teilweise geordnete Skala bildet das zweittiefste Meßniveau. Eine solche Skala liegt dann vor, wenn eine bestimmte Anzahl von Äquivalenzklassen vorhanden ist und die Beziehung "größer als" (Ordnung) nur für einige, nicht aber für alle Klassenpaare zutrifft.²⁾ Für diesen in den Sozialwissenschaften höchst relevanten Skalentyp bringt COOMBS das Beispiel der Messung des sozioökonomischen Status. Wenn der sozioökonomische Status mit Hilfe zweier Indikatoren, Einkommen und Ausbildungsniveau, gemessen wird, so bestehen folgende Möglichkeiten: Eine Klasse von Individuen A bezieht mehr Einkommen und hat gleichzeitig auch ein höheres Ausbildungsniveau als eine andere Klasse von Individuen B. Die Individuen der Klasse A haben dann einen höheren sozioökonomischen Status als die Individuen der Klasse B, und es gilt $A > B$. Wenn nun die Individuen der

1) BESSLER, H.: Zur Theoriebildung in der Massenkommunikationsforschung, a.a.O., S. 130

2) Vgl. COOMBS, C. H.: Theory and Methods of Social Measurement, in: L. FESTINGER und D. KATZ (Hrsg.): Research Methods in the Behavioral Sciences, New York, Chicago, San Francisco, Toronto und London 1953, S. 474 f.

Klasse B mehr Einkommen beziehen und ebenfalls ein höheres Ausbildungsniveau als die einer weiteren Klasse C haben, so gilt nicht nur $B > C$, sondern auch $A > C$ ¹⁾. Es gebe nun eine vierte Klasse von Individuen D, die weniger Einkommen und ein geringes Ausbildungsniveau als die der Klasse A haben, aber mehr Einkommen und ein höheres Ausbildungsniveau als die der Klasse C; dann gilt $A > D > C$. Haben aber die Individuen der Klasse D ein höheres Einkommen, jedoch gleichzeitig ein geringeres Ausbildungsniveau als die von B, dann kann nicht ohne weiteres entschieden werden, ob $B > D$ oder $D > B$ gilt, oder ob eine solche Beziehung ("größer als") überhaupt besteht. Das heißt, daß für die Klassen B und D kein Größenverhältnis angegeben werden kann; in einem solchen Fall spricht COOMBS von einer teilweise geordneten Skala. COOMBS glaubt, daß dieses Meßniveau wahrscheinlich dann erreicht wird, "whenever an attribute is made up of two or more primitive attributes which do not combine additively or compensate for each other"²⁾.

c) Die Ordinalskala

Bei einer Ordinalskala werden die Zahlen den verschiedenen Ausprägungen einer zu messenden Eigenschaft so zugeordnet, daß die Ordnung der Zahlen der Ordnung der Stärke verschiedener Ausprägungen entspricht. Zu den Postulaten der Nominalskala kommen nunmehr noch die Postulate der Rangfolge, die jetzt - anders als bei der teilweise geordneten Skala - für alle Ausprägungen der zu messenden Eigenschaften gelten:

-
- 1) Die Postulate der Rangfolge werden im nächsten Abschnitt über die Ordinalskala genau formuliert.
 - 2) COOMBS, C. H.: Theory and Methods of Social Measurement, a.a.O., S. 475

1. Wenn $i > j$ dann $j \succ i$.
2. Wenn $i > j$ und $j > k$ dann $i > k$

Als Rechenoperationen können auf dem Niveau der Ordinalskala ausgeführt werden: das Bilden von Rangreihen, die Berechnung von Median, Centilen und von Rangkoeffizienten. Nur solche Transformationen, die die Ordnung beibehalten (monotone Transformationen), sind möglich.

Ein typisches Beispiel einer Ordinalskala ist die MOHS'sche Härteskala: Gips (j) ritzt Talg (i), Kalkspat (k) ritzt Gips (j), usw. Lediglich die Ordnung der Zahlen hat in diesem Fall eine empirische Bedeutung, nicht dagegen die Kriterien des Abstandes und des Ursprunges. In dem oben skizzierten theoretischen Ansatz zur Erklärung der Entstehung und Wirkung von Aussagen der Massenkommunikation wird - wie bereits ausgeführt - von der Möglichkeit einer Messung von Einstellungen von Kommunikatoren und Rezipienten auf Ordinalskalenniveau ausgegangen, um den Ansatz einer empirischen Prüfung zugänglich machen zu können.¹⁾

d) Die Ordinalskala mit natürlichem Ursprung

Eine Ordinalskala mit natürlichem Ursprung hat die unter c) beschriebenen Charakteristika des Typs der Ordinalskala, jedoch mit dem Zusatz, daß die Zahlenreihe hier einen natürlichen Ursprung hat, dem die Zahl "null" zugeordnet wird.²⁾ Das bedeutet, daß bei diesem Skalentyp außer dem

1) Vgl. Abschnitt 242 dieser Arbeit.

2) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 16 und 19 f. Zum Problem des natürlichen Ursprunges bzw. des absoluten Nullpunktes vgl. unter g) die Ausführungen über die Ratioskala.

Kriterium der Ordnung auch noch das Kriterium des Ursprunges empirische Bedeutung erhält. Alle Transformationen, die den Ursprung und die Ordnung unverändert lassen (monotone Transformationen durch den Ursprung), sind möglich.

e) Die geordnete metrische Skala

Wie die Ordinalskala mit natürlichem Ursprung, so liegt auch die geordnete metrische Skala im Meßniveau zwischen den klassischen Typen der Ordinal- und der Intervallskala. Hier liegt ein Skalentyp vor, bei dem erstmals dem Kriterium des Abstandes empirische Bedeutung zukommt. Eine geordnete metrische Skala ist dann gegeben, wenn zwischen mindestens drei aufeinanderfolgenden Klassen (i , j und k) die Beziehung $i > j > k$ besteht und man zumindest für einige Abstände zwischen den Klassen (z.B. $\bar{ij}$, $\bar{jk}$) die Größenverhältnisse angeben kann, z.B. $\bar{ij} > \bar{jk}$ oder $\bar{jk} > \bar{ij}$.¹⁾ Es müssen also Aussagen über die Größenverhältnisse zwischen den Abständen e i n i g e r der vorhandenen Klassen möglich sein, damit man von einer geordneten metrischen Skala sprechen kann. Wenn man eine Aussage über die Größenverhältnisse der Abstände zwischen a l l e n vorhandenen Klassenpaaren machen kann, so liegt eine höher geordnete metrische Skala vor, die sich wiederum etwas stärker dem Intervallskalenniveau nähert. COOMBS glaubt, mit der von ihm entwickelten Technik der transferierten Einschätzungen bei der Messung von Einstellungen das Niveau einer geordneten metrischen Skala oder einer höher geordneten metrischen Skala erreichen zu können.²⁾ Über das Kriterium des

1) Vgl. COOMBS, C. H.: Theory and Methods of Social Measurement, a.a.O., S. 477 - 481; SIEGEL, S.: A Method for Obtaining an Ordered Metric Scale, Psychometrika 21, 1956, S. 207 - 216

2) Vgl. unten Abschnitt 435

Ursprunges wird bei der geordneten metrischen Skala im Gegensatz zur Ordinalskala mit natürlichem Ursprung nichts gesagt; man kann deshalb nicht ohne weiteres davon ausgehen, daß bei der geordneten metrischen Skala ein höheres Meßniveau vorliegt als bei der Ordinalskala mit natürlichem Ursprung.¹⁾

SIXTL vertritt die Ansicht, daß die Bildung von "Zwischentypen" - wie es z.B. die geordnete metrische Skala ist - gegenüber der von STEVENS vorgeschlagenen Einteilung in Nominal-, Ordinal-, Intervall- und Ratioskalen keine Vorteile bringe, weil als zulässige Rechenoperationen jeweils nur die des nächstunteren Meßniveaus in Frage kämen.²⁾ Es ist jedoch anzunehmen, daß viele der in der empirischen Sozialforschung entwickelten Skalen die Eigenschaften z.B. von teilweise geordneten oder geordneten metrischen Skalen aufweisen. Die Konstruktion und Weiterentwicklung von Rechenoperationen, die speziell auf diese Meßniveaus zugeschnitten sind, erscheinen deshalb durchaus sinnvoll.

f) Die Intervallskala

Bei einer Intervallskala entspricht die Ordnung der Zahlen wie bei einer Ordinalskala der Ordnung der Ausprägung der zu messenden Eigenschaft. Hinzu kommt noch, daß die Abstände zwischen den Zahlen empirische Bedeutung haben; das

1) Bei den Punkten a) bis g) dieses Abschnittes handelt es sich also eigentlich um eine teilweise geordnete Skala der Skalentypen. - Es ließen sich freilich auch geordnete metrische Skalen mit natürlichem Ursprung denken; solche Skalen würden dann tatsächlich ein höheres Meßniveau als Ordinalskalen mit natürlichem Ursprung bilden.

2) Vgl. SIXTL, F.: Meßmethoden in der Psychologie, a.a.O., S. 9

heißt, daß diese Abstände entsprechend unterschiedliche Ausprägungen der zu messenden Eigenschaft repräsentieren. Bei einer Intervallskala kommt also den Kriterien der Ordnung und des Abstandes empirische Bedeutung zu, nicht aber dem Kriterium des Ursprunges. Zu den für Nominal- und Ordinalskala gültigen Postulaten treten hier noch die Postulate der Additivität:

1. Wenn $i = k$ und $j > 0$ dann $i + j > k$
2. $i + j = j + i$. Man spricht hier von Kommutativität; das bedeutet, daß das Ergebnis einer Addition nicht durch die Reihenfolge, in der addiert wird, beeinflußt ist.
3. Wenn $i = k$ und $j = l$ dann gilt $i + j = k + l$. Dieses Postulat besagt, daß identische Objekte untereinander austauschbar sind.
4. $(i + j) + k = i + (j + k)$. Diese Beziehung nennt man assoziativ; das heißt, "wenn man ein beliebiges Paar von Summanden addiert und deren Summe zu den verbleibenden Gliedern addiert, ändert sich nichts am Ergebnis".¹⁾

Als zulässige Rechenoperationen kommen auf diesem Skalenniveau die Berechnung von arithmetischem Mittel, Standardabweichung und BRAVAIS-PEARSON'schen Korrelationskoeffizienten in Frage. Lineare Transformationen der Form $y = ax + b$ sind möglich. Typisches Beispiel für eine Intervallskala ist die Messung der Temperatur mit Hilfe einer Fahrenheit- oder Celsiuskala.

Gleichheit der Intervalle oder zumindest Bestimmbarkeit der Intervallbreiten ist für eine Skala eine durchaus wünschenswerte Eigenschaft, weil nur dann berechtigterweise höhere mathematische Operationen zur Analyse der Meßdaten

1) SIXTL, F.: Meßmethoden in der Psychologie, a.a.O., S. 5

verwendet werden können. Die auf THURSTONE zurückgehenden "Judgment"-Techniken sind die einzigen etablierten Vorgehensweisen, die eine Bestimmung der Intervallbreiten zu erreichen versuchen. Der Nachweis des Vorliegens gleicher Intervalle oder die Bestimmung der Intervallbreiten sind jedoch bisher für die in den Sozialwissenschaften verwendeten Skalen zur Messung von Einstellungen nicht überzeugend gelungen. Alle Einstellungsskalen müßten deswegen eigentlich wie Ordinalskalen oder bestenfalls wie geordnete metrische Skalen behandelt werden. Es ist deshalb nur sinnvoll, wenn ein theoretischer Ansatz zur Erklärung der Entstehung und Wirkung von Aussagen der Massenmedien, dessen zentrale Variablen Einstellungen sind, davon ausgeht, daß diese Einstellungen nur auf Ordinalskalenniveau gemessen werden können.

g) Die Ratioskala

Die Ratioskala als höchstes Meßniveau kann wie die unter f) beschriebene Intervallskala charakterisiert werden; sie weist jedoch das zusätzliche Merkmal auf, daß die den verschiedenen Ausprägungen der zu messenden Eigenschaft zugeordneten Zahlen dem Abstand dieser Ausprägungen von einem natürlichen Ursprung (absoluter Nullpunkt) entsprechen müssen. Damit erhält neben den Kriterien der Ordnung und des Abstandes auch das Kriterium des Ursprunges empirische Bedeutung. Das heißt, daß auch die Verhältnisse der Zahlen zueinander empirische Bedeutung gewinnen. Während sich auf dem Niveau einer Intervallskala nur Additionen bzw. Subtraktionen ausführen lassen, kann beim Vorliegen einer Ratioskala auch multipliziert bzw. dividiert werden. Lineare Transformationen der Form $y = ax$, die den Nullpunkt unverändert lassen, sind möglich. STEVENS spricht bei dieser

Form der Transformation (Multiplikation mit einer Konstante) von Ähnlichkeitstransformation ("similarity transformation").¹⁾ Typische Beispiele für Ratioskalen sind die Messung von Länge und Gewicht in der Physik.

Das Problem des natürlichen Ursprunges bzw. des absoluten Nullpunktes bereitet in den Sozialwissenschaften und vor allem bei der Messung von Einstellungen²⁾ besondere Schwierigkeiten. Die Festlegung eines absoluten Nullpunktes einer Skala ist nur mit Hilfe einer entsprechenden Theorie möglich. Ist ein solcher Nullpunkt festlegbar, so bedeutet das, daß Multiplikationen bzw. Divisionen der Meßergebnisse im Lichte der betreffenden Theorie eindeutig interpretiert werden können. Das ist aber bei den Skalen zur Messung von Einstellungen bisher noch nicht der Fall. Die Bestimmung eines absoluten Nullpunktes ist hier bis jetzt nicht möglich. Im Falle der Einstellungsmessung wird zumeist der Punkt auf einer Skala als Nullpunkt bezeichnet, an dem eine positive Einstellung in eine negative umschlägt oder umgekehrt. Dieser Punkt wird vielfach auch als "neutraler" Punkt bezeichnet.³⁾ Damit wird jedoch eine völlig andere Art von Nullpunkt konstruiert, als er z.B. bei der Längen- oder Gewichtsmessung in der Physik vorliegt. Das Problem scheint weder theoretisch noch praktisch zufriedenstellend gelöst zu sein⁴⁾; das zeigt sich

1) Vgl. STEVENS, S. S.: Mathematics, Measurement, and Psychophysics, a.a.O., S. 24

2) Diese Frage wird auch im Zusammenhang mit der Einstellungsmessung gelegentlich erörtert, obwohl es sich bei den Skalen zur Messung von Einstellungen doch bestenfalls um geordnete metrische Skalen handelt und sich damit das Problem des absoluten Nullpunktes zunächst gar nicht stellt.

3) Vgl. z.B. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 30 f. Hier wird das Problem anhand der Ordinalskala mit natürlichem Ursprung behandelt.

4) Vgl. SHAW, M. E. und J. M. WRIGHT: Scales for the Measurement of Attitudes, New York, St. Louis, San Francisco, Toronto, London und Sidney 1967, S. 21

nicht zuletzt an der ausgesprochenen Unklarheit der entsprechenden Stellen in der Literatur.

33 Arten der Messung

Die Unterscheidung zwischen den verschiedenen Skalentypen hinsichtlich des Niveaus der Messung beruht darauf, wie viele Informationen über die zu messende Eigenschaft die einzelnen Skalen enthalten. Im folgenden ist nun zwischen verschiedenen Arten von Informationen zu unterscheiden. Das heißt, es stellt sich die Frage, auf welche Weise die Kriterien der Ordnung, des Abstandes und des Ursprunges empirische Bedeutung erhalten können.¹⁾ Dabei lassen sich drei verschiedene Arten der Messung unterscheiden, die zunächst anhand der schematischen Darstellung in Abschnitt 23 dieser Arbeit²⁾ kurz erläutert und anschließend ausführlicher betrachtet werden sollen:

- a) Bei der Messung per definitionem³⁾ werden die Beziehungen zwischen den K'- und den K-Konstrukten nicht geprüft, sondern es wird a priori eine Übereinstimmung zwischen diesen beiden Arten von Konstrukten angenommen.
- b) Bei der fundamentalen oder grundlegenden Messung bemüht man sich dagegen darum, die Beziehungen zwischen den K'- und den K-Konstrukten auf verschiedenen noch darzustellenden Wegen zu klären.

1) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 21

2) Vgl. oben Figur 1, S. 17

3) TORGERSON spricht von "measurement by fiat", vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 31 f. In der Literatur findet sich auch der Ausdruck "measurement by definition". Vgl. z.B. GOUDE, G.: On Fundamental Measurement in Psychology, Stockholm, Göteborg und Uppsala 1962, S. 15 f. Wenn GOUDE von "measurement by definition" spricht, meint er damit jedoch, wie das von ihm angeführte Beispiel zeigt, etwas anderes als hier unter Messung per definitionem verstanden werden soll.

c) Während die Konstrukte K_1 bis K_4 in der Figur 1 fundamental gemessen werden könnten, indem man die Beziehungen zwischen den Konstrukten K'_1 bis K'_4 zu den entsprechenden Konstrukten K_1 bis K_4 zu klären versucht, ist das bei den Konstrukten K_5 und K_6 nicht möglich. Die Messung dieser Konstrukte setzt vielmehr die Messung anderer Konstrukte voraus; sie können folglich keine unmittelbare operationale Bedeutung erlangen. In diesem Falle spricht man von *abgeleiteter Messung*.

331 Messung per definitionem

Bereits die kurzen Ausführungen zu Beginn dieses Abschnittes zeigen, daß eine Erörterung der Arten der Messung unter anderem auf das Validitätsproblem hinausläuft: Bei der Messung per definitionem werden die Beziehungen zwischen konstitutiver und operationaler Bedeutung eines Konstruktes nicht untersucht. Um diese Art der Messung handelt es sich bei den in den Sozialwissenschaften verwendeten Indikatoren jedoch fast immer.¹⁾ Diese Indikatoren besitzen - zumindest im Anfangsstadium der Entwicklung einer Theorie - ausschließlich operationale Bedeutung, wohingegen die konstitutive Bedeutung fehlt. Häufig zitiertes Beispiel in diesem Zusammenhang ist die Messung der Intelligenz und die charakteristische Formulierung: Intelligenz ist, was der Intelligenztest mißt.

Obwohl an dieser Art der Messung, wie TORGERSON betont, nichts falsch oder logisch inkorrekt ist, erscheint sie doch im Hinblick auf den Fortschritt der Theorienbildung

1) Zu den folgenden Überlegungen vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 22 und 23 ff.

überaus problematisch: Beobachtbare Eigenschaften werden ausgewählt und a priori als mit dem in der Theorie verwendeten Konstrukt in Übereinstimmung stehend angesehen. Eine Messung dieser beobachtbaren Eigenschaften wird als Messung des theoretischen Konstruktes betrachtet, und die auf diese Weise gewonnenen Werte werden oftmals auch noch so behandelt, als ob es sich um Intervall- oder Ratio-skalen handle. Es ist klar, daß bei der Messung per definitionem die tatsäc h l i c h e Übereinstimmung von konstitutiver und operationaler Bedeutung eines Konstruktes besonders fraglich ist. Die eigentliche Schwierigkeit besteht darin, daß ein Konstrukt auf viele unterschiedliche Weisen operationalisiert werden kann. In empirischen Untersuchungen können dann unterschiedliche Operationalisierungen zu übereinstimmenden oder zu divergierenden Ergebnissen führen, wobei im Falle divergierender Ergebnisse den operationalen Definitionen verschiedene Konstrukte zugrunde gelegen haben.¹⁾

Die Unklarheit bezüglich der Übereinstimmung von konstitutiver und operationaler Bedeutung der in einer Hypothese verwendeten Konstrukte gerade bei dieser Art der Messung führt - wie bereits erwähnt - zu der relativen Immunität von Theorien gegenüber widersprechenden Ergebnissen. Es ist daher TORGERSON zuzustimmen, wenn er fordert, daß es beim gegenwärtigen Stand und im Hinblick auf die zukünftige Entwicklung der Sozialwissenschaften besonders wichtig erscheint, sich mit der Entwicklung und Prüfung von Meßtechniken zu beschäftigen, mit deren Hilfe relevante theoretische Konstrukte auf fundamentale Weise gemessen werden können, um so zumindest einen Teil der im Meßvorgang liegenden Unsicherheit abbauen zu können.

1) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 25

332 Grundlegende Messung (fundamental measurement)

Man nennt eine Messung dann grundlegend, wenn sie zur Erfassung einer bestimmten Eigenschaft nicht die Messung einer oder mehrerer anderer Eigenschaften voraussetzt.¹⁾ Das trifft, wie bereits ausgeführt, auf die Konstrukte K_1 bis K_4 in der schematischen Darstellung des Abschnittes 23 zu. Hinzu kommt noch ein zweites Charakteristikum: Grundlegende Messung beruht darauf, daß die verschiedenen Ausprägungen eines bestimmten Konstruktes durch Gesetze (natural laws) miteinander in Beziehung stehen.²⁾ In diesem Sinne beschreibt GOUDE grundlegende Messung:

- a) "by the empirical interpretation - with the aid of an instrument of measurement, and in accordance with certain rules - of certain of the numbers, signs of relation (or operation), axioms and theorems of mathematics, and
- b) by the fact that it does not use (or presuppose) any other previous measurement".³⁾

Grundlegende Messung beruht unmittelbar darauf, daß es bestimmte strukturelle Ähnlichkeiten zwischen den mathematischen Operationen, die bei einer Messung angewendet werden, und der Realität der Objekte, deren Eigenschaft gemessen werden sollen, gibt.⁴⁾ Die Kriterien der Ordnung, des Abstandes und des Ursprunges können durch diese Ähnlichkeiten eine, wenn auch unvollständige empirische Be-

1) Vgl. CAMPBELL, N. R.: An Account of the Principles of Measurements and Calculations, London 1928, S. 14. CAMPBELL spricht jedoch von Messung nur beim Vorliegen von Intervall- oder Ratioskalen.

2) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 22

3) GOUDE, G.: On Fundamental Measurement in Psychology, a.a.O., S. 15

4) Vgl. ebenda

deutung erhalten, und zwar unabhängig von Gesetzen bzw. Theorien, die die zu messende Eigenschaft mit Eigenschaften an d e r e r Variablen verbinden. Trotzdem kann ein so gemessenes Konstrukt natürlich mit anderen Konstrukten in Beziehung stehen; solche Beziehungen müssen sogar vorhanden sein, wenn das betreffende Konstrukt überhaupt theoretische Relevanz haben soll. Ein auf diese Art gemessenes Konstrukt besitzt sowohl operationale als auch konstitutive Bedeutung, weil einerseits definitive Regeln für die Zuordnung der Zahlen bestehen, um die verschiedenen Ausprägungen abzubilden, und weil die Zahlen andererseits Gesetze (natural laws) widerspiegeln, die die verschiedenen Ausprägungen der betreffenden Eigenschaft verbinden.¹⁾

Ein Beispiel grundlegender messung auf Ratioskalenniveau ist in den Naturwissenschaften die Messung der Länge als eines Konstruktes von großer theoretischer Relevanz. Die Längenmessung setzt die Messung einer anderen Variablen nicht voraus, sie kann unmittelbar erfolgen. Zudem stehen die verschiedenen Ausprägungen durch Gesetze miteinander in Beziehung: Wenn zwei Stäbe bestimmter Länge mit ihren Enden aneinandergelegt werden, so entsteht eine Kombination (Addition) der Ausprägungen der Eigenschaft "Länge" beider Stäbe. Einem der beiden Stäbe kann eine Zahl willkürlich zugeordnet werden, um seine Länge "abzubilden"; ist das geschehen, so sind damit die Zahlen für alle anderen Ausprägungen, z.B. auch für die Länge des zweiten Stabes, festgelegt.²⁾

1) Vgl. zu diesen Ausführungen TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 22

2) Die Logik dieser Vorgehensweisen soll hier nicht im einzelnen beschrieben werden. Vgl. dazu HEMPEL, C. G.: Fundamentals of Concept Formation in Empirical Science, a.a.O., S. 62 - 67; TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 33 f.; SUPPES, P. und J. L. ZINNES: Basic Measurement Theory, a.a.O., S. 4 - 17 und 22 - 48

Die im Kapitel 4 dieser Arbeit erläuterten Techniken stellen Versuche der grundlegenden Messung von Einstellungen dar. In diesem Zusammenhang ist noch einmal darauf hinzuweisen, daß auch im Falle grundlegender Messung die Beziehungen zwischen konstitutiver und operationaler Bedeutung eines in einer Hypothese verwendeten Konstruktes nicht endgültig geklärt werden können: "It would therefore be quite unreasonable to demand of a fundamental measurement that the empirical interpretations should be exact or exhaustive. What we can demand is that the interpretations made shall be natural, directly relevant and sufficiently accurate for the context in which the measures are to be used."¹⁾ Um den in diesem Zitat gestellten Anforderungen genügen zu können, muß ein Meßinstrument strengen Reliabilitäts- und Validitätsprüfungen unterzogen werden, wenn es ein Konstrukt auf dem Wege grundlegender Messung erfassen soll. Validitätsüberlegungen spielen nur im Rahmen der grundlegenden Messung eine unmittelbare Rolle. Bei der Messung per definitionem wird diese Frage, wie bereits ausgeführt, weitgehend ignoriert. Abgeleitete Messung eines Konstruktes setzt, wie noch zu zeigen ist, die reliable und valide grundlegende Messung zumindest eines anderen Konstruktes voraus; Validitätsüberlegungen spielen in diesem Falle nur mittelbar eine Rolle.

1) GOUDE, G.: On Fundamental Measurement in Psychology, a.a.O., S. 15

3321 Die Bestimmung der Reliabilität¹⁾

Reliabilität bedeutet Zuverlässigkeit: Eine Messung ist dann reliabel (zuverlässig), wenn sie keine zufälligen Fehler aufweist. Validität bedeutet Gültigkeit: Eine Messung ist dann valide (gültig), wenn sie keine systematischen Fehler aufweist. So jedenfalls lauten herkömmlicherweise die Definitionen von Reliabilität und Validität. Die Diskussion der Reliabilität einer Messung bezieht sich also auf die mögliche unsystematische Variation in den Meßergebnissen und deren Aufdeckung bzw. Behebung. Reliabilität kann demnach definiert werden als das Ausmaß, zu dem eine Skala zu konsistenten Meßergebnissen führt, wenn sie mehrmals zur Messung einer gleichbleibenden Eigenschaft eingesetzt wird. GHISELLI definiert etwas abweichend hiervon Reliabilität als "the extent of unsystematic variation in the quantitative descriptions of the amount of some trait an individual possesses or manifests when that trait is measured a number of times."²⁾ Wenn man die Reliabilität eines Meßinstrumentes bestimmen will, versucht man demnach herauszufinden, in welchem Ausmaß die Ergebnisse, die mit Hilfe des betreffenden Instrumentes erzielt werden, durch zufällige Störfaktoren beeinflusst sind.

1) Das Reliabilitätsproblem wird hier im Rahmen der grundlegenden Messung behandelt, obwohl es an sich auch bei der Messung per definitionem relevant ist. Da es sich jedoch bei den in Kapitel 4 erläuterten Skalierungstechniken um Versuche grundlegender Messung handelt, erscheint eine Behandlung an dieser Stelle eher angebracht. Auf die in diesem Abschnitt zu erörternden Fragen soll möglichst kurz eingegangen werden, weil hierüber einerseits alle einschlägigen Lehrbücher berichten. Vgl. z.B. SELLITZ, C., M. JAHODA, M. DEUTSCH und S. W. COOK: Research Methods in Social Relations, a.a.O., S. 166 - 186.

Zum anderen ist das Reliabilitätsproblem eher von "praktischer" als von theoretischer Relevanz; und die vorliegende Arbeit möchte hauptsächlich einen Beitrag zur Klärung des Verhältnisses von Theorienbildung und Messung bei der Verwendung des Konstruktes "Einstellung" leisten.

2) GHISELLI, E. E.: Theory of Psychological Measurement, New York, San Francisco, Toronto und London 1964, S. 215

Reliabilität ist eine Eigenschaft von Skalen, die im Vergleich zur Validität noch relativ einfach erreichbar zu sein scheint. Jedenfalls muß man diesen Eindruck gewinnen, wenn man sich in der Sammlung verschiedenster Einstellungsskalen von SHAW und WRIGHT¹⁾ die den einzelnen Skalen vorangestellten Bemerkungen über Validität und Reliabilität ansieht: Während die Reliabilität oft recht gut mittels der verschiedenen zur Verfügung stehenden Vorgehensweisen bestimmbar ist, kann über die Validität dieser Skalen meist nur sehr Vages und Unzureichendes berichtet werden.

In der Literatur werden zumeist drei solcher Vorgehensweisen genannt, die es ermöglichen, die Reliabilität einer Skala empirisch zu bestimmen: das Wiederholungsverfahren, das Verfahren äquivalenter Formen und das Halbierungsverfahren.²⁾

- a) Das W i e d e r h o l u n g s v e r f a h r e n (test-retest method): Bei der Anwendung des Wiederholungsverfahrens wird dieselbe Skala bei einer bestimmten Gruppe von Versuchspersonen zwei-oder mehrmals eingesetzt. Ein Korrelationskoeffizient wird als Maß für die Reliabilität benutzt; dabei werden die von den Versuchspersonen bei den einzelnen, zeitlich auseinanderfallenden Anwendungen der Skala erreichten Punktwerte korreliert.

-
- 1) Vgl. SHAW, M. E. und J. M. WRIGHT: Scales for the Measurement of Attitudes, a.a.O.
- 2) Die Darstellungen der drei Verfahren folgt im wesentlichen SHAW, M. E. und J. M. WRIGHT: Scales for the Measurement of Attitudes, a.a.O., S. 16 f., die eine sehr knappe und brauchbare Übersicht der einzelnen Vorgehensweisen und der damit verbundenen Vor- und Nachteile geben. Zur Reliabilität bei Techniken der Aussagenanalyse vgl. BESSLER, H.: Zur Theoriebildung in der Massenkommunikationsforschung, a.a.O., S. 135 ff.

Dieses Verfahren hat im Vergleich zu den anderen den Vorteil, daß immer dieselben Items benutzt werden können, und Unterschiede zwischen den Items auf diese Weise als Grund für mangelnde Reliabilität ausschließen. Zudem wird jeweils nur eine Skala benötigt, was unter forschungsökonomischen Gesichtspunkten von Bedeutung ist. Auf der anderen Seite hat das Wiederholungsverfahren erhebliche Nachteile: Der wichtigste Mangel des Verfahrens besteht darin, daß die erste Anwendung eines Instrumentes bei einer Gruppe von Versuchspersonen die nachfolgenden beeinflussen kann; solche Beeinflussungen werden "carry-over effects"¹⁾ genannt. Sie können dazu führen, daß die Reliabilität eines Instrumentes zu hoch veranschlagt wird. Das dürfte vor allem dann der Fall sein, wenn sich die Versuchspersonen bei der zweiten Messung noch an bestimmte Items der ersten Messung erinnern und zu diesen Items einfach wie beim ersten Mal Stellung nehmen. Andererseits können die "carry-over effects" jedoch auch ein Grund dafür sein, die Reliabilität eines Instruments niedriger einzuschätzen, als es tatsächlich der Fall ist. Das wird dann vorkommen, wenn die erste Anwendung eines Instrumentes die Einstellungen einer Versuchsperson geändert hat. Außerdem können beim Wiederholungsverfahren zwischen der ersten und den weiteren Messungen einstellungsändernde Ereignisse stattgefunden haben, die sich auf die Reliabilitätsfeststellung negativ auswirken.

Diese Nachteile des Wiederholungsverfahrens können nicht mit Sicherheit ausgeschaltet werden; sie können, wie SHAW und WRIGHT feststellen, nur dadurch in ihren negativen Auswirkungen eingeschränkt werden, daß man zwit-

1) Vgl. GHISELLI, E. E.: *Theory of Psychological Measurement*, a.a.O., S. 277 f.

schen der ersten und den weiteren Messungen einen Abstand wählt, der einerseits lang genug ist, um die Einflüsse des Gedächtnisses auszuschalten, und der andererseits kurz genug ist, um die störenden Einflüsse anderer Variablen herabzumindern. Unter diesen Gesichtspunkten werden Abstände zwischen zwei und sechs Wochen empfohlen.

- b) Das Verfahren äquivalenter Formen (equivalent forms method): Bei diesem Verfahren werden den Versuchspersonen zwei oder mehr Skalen vorgelegt, die als äquivalent in bezug auf die zu messende Eigenschaft angesehen werden. Durch Korrelation der jeweils von den Versuchspersonen erreichten Punktwerte wird die Reliabilität bestimmt.

Diese Vorgehensweise hat den Vorteil, daß die möglichen negativen Auswirkungen des zeitlichen Abstandes zwischen den Messungen minimal sind. Eine zu hohe Reliabilität kann jedoch dann errechnet werden, wenn sich bei den Versuchspersonen bestimmte Antwortmuster ("response patterns") herausbilden. Der Hauptnachteil des Verfahrens besteht nach der zusammenfassenden Beurteilung von SHAW und WRIGHT darin, daß - anders als beim Wiederholungsverfahren - zumindest zwei Formen einer Skala benötigt werden und daß ein errechneter Korrelationskoeffizient nicht nur die Reliabilität der Messung zum Ausdruck bringt, sondern auch ein Maß dafür ist, wieweit die einzelnen Skalenformen tatsächlich dasselbe messen. Die Vorteile des Wiederholungsverfahrens sind demnach im wesentlichen mit den Nachteilen des Verfahrens äquivalenter Formen identisch, und die Vorteile des Verfahrens äquivalenter Formen sind gleichzeitig die Nachteile des Wiederholungsverfahrens.

- c) Das Halbierungsverfahren (split-half method): Zur Bestimmung der Reliabilität wird hier eine Skala in zwei (oder mehr) Teile aufgespalten; die einzelnen Teile werden dann wie selbständige Skalen behandelt. Als Maß der Reliabilität dient die Korrelation zwischen den einzelnen Skalenteilen; zur Berechnung kann die SPEARMAN-BROWN-Formel verwendet werden.¹⁾ Die Teilung einer Skala kann soweit fortgeführt werden, bis man im Extremfall zu den einzelnen Items gelangt; diese Vorgehensweise wird von GHISELLI als selbständige Form zur empirischen Bestimmung der Reliabilität angeführt.²⁾

Das Halbierungsverfahren hat dieselben Vorteile wie die Technik der äquivalenten Formen; zudem wird hier nur eine Skala benötigt. Eigentlich handelt es sich jedoch bei dieser Vorgehensweise um die Messung der internen Konsistenz der Items bei einmaliger Anwendung einer Skala, nicht aber um die Messung der Konsistenz bei wiederholten Anwendungen eines Instrumentes.

Die genannten Verfahren führen zu einer mehr oder weniger genauen Bestimmung der Reliabilität, wobei die Entscheidung darüber, ob man es nun mit einem reliablen Instrument zu tun hat oder nicht, recht willkürlich getroffen werden muß. So kann man z.B. durch Konvention einen bestimmten Korrelationsgrad als Minimum einführen. Es gibt zwar verschiedene Möglichkeiten zur Verbesserung der Reliabilität einer Skala³⁾, vollständige Reliabilität ist

-
- 1) Vgl. z.B. die Diskussion der Formel bei SELTZ, C., M. JAHODA, M. DEUTSCH und S. W. COOK: Research Methods in Social Relations, a.a.O., S. 183; dort auch weitere Literatur.
- 2) Vgl. GHISELLI, E. E.: Theory of Psychological Measurement, a.a.O., S. 284 ff.
- 3) Vgl. SELTZ, C., M. JAHODA, M. DEUTSCH und S. W. COOK: Research Methods in Social Relations, a.a.O., S. 182 ff. Vgl. GHISELLI, E. E.: Theory of Psychological Measurement, a.a.O., S. 288 ff.

jedoch nicht zu erreichen. Im praktischen Forschungsbetrieb kommt als erschwerender Umstand noch hinzu, daß sich die Reliabilität eines Meßinstrumentes umgekehrt proportional zu seiner Empfindlichkeit ("sensitivity") verhält. Unter Empfindlichkeit ist dabei die Diskriminationsfähigkeit ("discriminating power") eines Instrumentes zu verstehen.¹⁾ Es muß noch erwähnt werden, daß nicht alle hier dargestellten Verfahren zur Prüfung der Reliabilität für jede der in Kapitel 4 diskutierten Einstellungsskalen gleich gut geeignet sind; hierauf wird an den entsprechenden Stellen des folgenden Kapitels hingewiesen.

3322 Die Bestimmung der Validität

Die Ausführungen zum Validitätsproblem sind in der Literatur zum Teil höchst unklar; hinzu kommen noch terminologische Unterschiede, die das Verständnis erschweren. Gemäß der in dieser Arbeit verwendeten Terminologie handelt es sich bei der Erörterung der Validität von Meßinstrumenten, wie bereits ausgeführt, um die Frage: Wieweit stimmen Konstrukte in ihrer konstitutiven und in ihrer operationalen Bedeutung überein? Da das Validitätsproblem in seiner grundsätzlichen Bedeutung für Theoriebildung und -prüfung schon an anderer Stelle behandelt wurde,²⁾ kann in diesem Abschnitt eine Beschränkung derart erfolgen, daß lediglich eine kurze Übersicht der verschiedenen Möglichkeiten zur Prüfung der Validität von Einstellungsskalen gegeben wird.

1) Vgl. KAPLAN, A.: The Conduct of Inquiry, a.a.O., S. 200

2) Vgl. besonders Abschnitt 23

Herkömmlicherweise werden vier verschiedene Wege der Validitätsbestimmung genannt: "content validity", "predictive validity", "concurrent validity" und "construct validity"¹⁾. Da es sich jedoch bei der "predictive validity" und der "concurrent validity", wie zu zeigen ist, um zwei Unterformen derselben Vorgehensweise handelt, die hier als "criterion validity"²⁾ bezeichnet werden soll, werden dementsprechend im folgenden nur drei Arten der Validierung unterschieden. "Criterion" und "content validity" beziehen sich auf die Frage: "To what extent do the operations measure the traits they are supposed to measure?"³⁾ Das entspricht genau der bereits erwähnten Definition von Validität. Etwas abweichend davon bezieht sich die "construct validity" eher auf die Frage: "What traits are measured by the operations?"⁴⁾

- a) C o n t e n t - v a l i d i t y : Bei der "content validity", manche Autoren sprechen auch von "face validity", handelt es sich nicht eigentlich um eine strenge empirische Prüfung der Validität eines Instrumentes. Die Prüfung der "content validity" erfolgt durch

-
- 1) Vgl. AMERICAN PSYCHOLOGICAL ASSOCIATION COMMITTEE (Hrsg.): Technical Recommendations for Psychological Tests and Diagnostic Techniques, in: Psychological Bulletin Supplement 51, 1954, S. 1 ff.
Vgl. CRONBACH, L. J.: Essentials of Psychological Testing, New York, Evanston, London und Tokyo 1961, 2. Aufl., S. 96 ff., besonders die Übersicht S. 106, (1. Aufl. 1949).
Die englischen Ausdrücke werden hier beibehalten, weil die deutschen Übersetzungen zum Teil mißverständlich sind.
- 2) Vgl. PHILLIPS, B. S.: Social Research, New York und London 1960, S. 160 f.
- 3) GHISELLI, E. E.: Theory of Psychological Measurement, a.a.O., S. 337.
GHISELLI benutzt zwar den Terminus "predictive validity"; die Frage bezieht sich jedoch auch auf "criterion validity" als deren Unterform die "predictive validity" bezeichnet wurde.
- 4) GHISELLI, E. E.: Theory of Psychological Measurement, a.a.O., S. 337

die Abgabe von Expertenurteilen in verbaler Form; dabei wird der Grad, zu dem ein Meßinstrument die Eigenschaft mißt, die gemessen werden soll, anhand der Charakteristika oder des Inhalts des betreffenden Instrumentes und seiner Items eingeschätzt: "We examine a test or other measuring instrument and then, on the basis of our 'insight' or 'wisdom' as psychologists, judge its validity"¹⁾ Bei der Prüfung der "content validity" sind eigentlich zwei Fragen zu beantworten: (1) In welchem Maße bezieht sich jedes Element²⁾ des zu beurteilenden Instrumentes auf die zu messende Eigenschaft? (2) In welchem Maße repräsentiert das gesamte Instrument alle Aspekte der zu messenden Eigenschaft?³⁾ Die Möglichkeiten, diese Fragen zu beantworten, hängen sehr stark davon ab, wie genau die zu messenden Konstrukte konstitutiv definiert sind; denn je ungenauer und unvollständiger diese Definitionen sind, um so schlechter läßt sich z.B. beurteilen, wie weit die Items eines Meßinstrumentes eine repräsentative Auswahl aller Aspekte der zu messenden Eigenschaft bilden,⁴⁾ eine Frage, die gerade bei der Diskussion der Skalogramm-Analyse explizit eine Rolle spielt.

Die Bestimmung der "content validity" ist im Grunde wohl die unsicherste Form der Validierung eines Meßinstrumentes, einmal weil die konstitutive Bedeutung der bei der Theoriebildung in den Sozialwissenschaften verwendeten Konstrukte meist recht ungenau ist und zum an-

-
- 1) GHISELLI, E. E.: Theory of Psychological Measurement, a.a.O., S. 341
 - 2) Die Elemente der in dieser Arbeit diskutierten Instrumente zur Einstellungsmessung sind Items. Vgl. dazu Abschnitt 41
 - 3) Vgl. GHISELLI, E. E.: Theory of Psychological Measurement, a.a.O., S. 342
 - 4) Vgl. ders., a.a.O., S. 344

deren, weil sie eigentlich nur auf der Einsicht oder den Kenntnissen dessen beruht, der ein Urteil über die Validität abgibt: "Content validity in and of itself depends upon subjective judgment alone."¹⁾ Für diese Vorgehensweise spricht jedoch, daß die "content validity" gerade bei Einstellungsskalen leichter bestimmbar zu sein scheint als andere Formen der Validität; das zeigen die Bemerkungen über Validität, die den einzelnen Skalen in der Sammlung von SHAW und WRIGHT vorangestellt sind.²⁾

- b) **C r i t e r i o n - v a l i d i t y** : Bei der Prüfung der "criterion-validity" wird eine empirische Gültigkeitskontrolle eines Meßinstrumentes vorgenommen. Man geht dabei von der Frage aus: In welchem Ausmaß kann ein Meßinstrument durch ein anderes ersetzt werden?³⁾ - Zur Beantwortung dieser Frage benötigt man die Messung einer Eigenschaft, die als "perfekt" unterstellt wird ("criterion"). Es wird dann versucht, die einzelnen Ausprägungen bei dieser Messung mit Hilfe von Ergebnissen bei der Anwendung eines anderen Meßinstrumentes, das auf seine Validität geprüft werden soll ("predictor"), vorauszusagen. In der Regel werden Korrelationskoeffizienten als Maß der Beziehungen zwischen "predictor"- und "criterion"-Ausprägungen verwendet.

1) GHISELLI, E. E.: Theory of Psychological Measurement, a.a.O., S. 346

2) Vgl. SHAW, M. E. und J. M. WRIGHT: Scales for the Measurement of Attitudes, a.a.O.

3) Die folgenden Ausführungen über "criterion validity" beruhen weitgehend auf der Diskussion in dem bereits mehrfach zitierten Standardwerk von GHISELLI, E. E.: Theory of Psychological Measurement, a.a.O., S. 338 ff.

Bei der Prüfung der "criterion-validity" braucht man also außer dem Instrument, das auf seine Validität bei der Messung eines bestimmten Konstruktes geprüft werden soll, noch ein zweites Instrument zur Messung desselben Konstruktes. Validität und Reliabilität dieses zweiten Instrumentes werden als gegeben angesehen. Durch einen Vergleich der Meßergebnisse bei der Anwendung beider Instrumente kann dann die Validität des zu prüfenden Instrumentes abgeschätzt werden.

Je nach der Zeitabfolge der beiden Messungen lassen sich verschiedene Arten der "criterion-validity" unterscheiden: Werden die beiden Messungen gleichzeitig durchgeführt, so spricht man im allgemeinen von "concurrent validity". Die Anwendung des Instrumentes, das auf seine Validität geprüft werden soll, kann aber auch zeitlich vor oder nach der Anwendung des zweiten Instrumentes erfolgen. Nur in einem solchen Falle dürfte es - wenn überhaupt - angebracht sein, von "predictive-validity" zu reden. Die "concurrent-validity" und die beiden Formen der "predictive validity" stellen sich dann als zwei Unterformen der "criterion-validity" dar.¹⁾

GHISELLI bezeichnet im Gegensatz zu der hier vertretenen Auffassung diese Arten der empirischen Validitätskontrollen - gleichzeitige, vorangehende und nachfolgende Kontrollmessung - als Unterfälle der "predictive validity".²⁾ BESSLER betont jedoch zu Recht, daß der Terminus "predictive validity" in diesem Zusammenhang mißverständlich sein kann: "Mit den Messungen zur Bestimmung der Prognosegültigkeit werden keine wissenschaftlichen Voraussagen geprüft, sondern lediglich Vergleiche durchgeführt".³⁾ Die

1) Vgl. GHISELLI, E. E.: Theory of Psychological Measurement, a.a.O., S. 338 ff.

2) Vgl. ders., a.a.O., S. 340 f.

3) BESSLER, H.: Zur Theoriebildung in der Massenkommunikationsforschung, a.a.O., S. 145

Verwendung der Ausdrücke 'criterion validity' oder 'pragmatic validity'¹⁾ wäre demnach vorzuziehen. Da es sich hier jedoch lediglich um terminologische Schwierigkeiten handelt, sollte man dieser Frage kein allzu großes Gewicht beimessen.

Der wesentliche Nachteil des Verfahrens der "criterion validity" besteht darin, daß es mit einer unsicheren Annahme arbeitet: Zur Prüfung eines Instrumentes, dessen Validität kontrolliert werden soll, wird unterstellt, daß ein anderes Instrument, das zum Vergleich in der beschriebenen Weise herangezogen werden muß, valide und reliabel mißt.²⁾ Die Berechtigung einer solchen Unterstellung kann jedoch nicht nachgewiesen werden, da auch die Validität dieses zweiten Instrumentes wieder nur mit einem der hier beschriebenen Verfahren geprüft werden kann und somit auch auf unsicheren Annahmen beruht. Mit Hilfe dieser Vorgehensweise läßt sich also lediglich Evidenz für die Validität eines Instrumentes sammeln, ein endgültiger Nachweis kann nicht erbracht werden.

- c) **C o n s t r u c t - v a l i d i t y** ³⁾: Bei der "construct-validity" handelt es sich um ein Problem von besonderer theoretischer Relevanz: Nur im Zusammenhang mit einer mehr oder weniger gut ausgebauten Theorie ist die Bestimmung der "construct-validity" möglich. Die Prüfung eines Instrumentes auf "construct validity" bedeutet zugleich auch immer die Prüfung zumindest einer Hypothese innerhalb einer Theorie. Im Fehlen präziser Hypothesen ist der Grund dafür zu sehen,

-
- 1) Vgl. SELLTIZ, C., M. JAHODA, N. DEUSSCH und S. W. COOK: Research Methods in Social Relations, a.a.O., S. 157 f.
2) Vgl. hierzu auch BESSLER, H.: Zur Theoriebildung in der Massenkommunikationsforschung, a.a.O., S. 146
3) Vgl. hierzu vor allem den bekannten Aufsatz von CRONBACH, L. J. und P. E. MEEHL: Construct Validity in Psychological Tests, Psychological bulletin 52, 1955, S. 281 - 302

daß diese Art der Validierung in den Sozialwissenschaften und so auch in der Massenkommunikationsforschung bisher kaum angewendet wurde und wohl auch nicht anwendbar war. Erst mit einem weiteren Fortschritt in der Theoriebildung und -prüfung wird dieses Verfahren zunehmend an Bedeutung gewinnen.

Die Vorgehensweise bei der "construct-validity" kann am verständlichsten anhand der schematischen Darstellung im Abschnitt 23 dieser Arbeit erklärt werden:¹⁾

Das Verfahren setzt zur Prüfung der Validität eines Instrumentes bei der Messung eines bestimmten Konstruktes (z.B. K_2) eine oder mehrere bestätigte Hypothesen voraus, die dieses Konstrukt (K_2) mit einem oder mehreren anderen Konstrukten (z.B. K_1 , K_3 und K_4) in Beziehung setzen. Wenn nun vorausgesetzt wird, daß die letztgenannten Konstrukte (K_1 , K_3 und K_4) valide und reliabel gemessen werden können, so läßt sich aufgrund der Hypothese oder der Hypothesen, die K_2 mit K_1 , K_3 und K_4 verbinden, das Meßergebnis voraussagen, das zustande kommen müßte, wenn das zu prüfende Instrument K_2 tatsächlich valide mißt.

Daraus geht hervor, daß eine solche Vorgehensweise nur im Zusammenhang mit einer Theorie anwendbar ist: Diese Theorie, deren Konstrukte untereinander in Verbindung stehen, ermöglicht es, bestimmte Beziehungen zwischen den einzelnen Ausprägungen bei der Messung verschiedener Konstrukte anzunehmen. Das kann am Beispiel der Validierung von Einstellungsskalen näher erläutert werden²⁾: In

1) Vgl. Figur 1, S. 17

2) Vgl. SHAW, M. E. und J. M. WRIGHT: Scales for the Measurement of Attitudes, a.a.O., S. 19

diesem Falle wird die "construct validity" durch die Feststellung der Beziehungen zwischen einer bestimmten Position auf einer Einstellungsskala und anderen Aspekten der Persönlichkeit, z.B. auch anderen Einstellungen, als sie die zu prüfende Skala mißt, ermittelt. Dabei sind es die theoretischen Vorstellungen, die für die betreffende Einstellung als Konstrukt mit konstitutiver Bedeutung bestehen, die verschiedene Arten und Ausmaße von Beziehungen zwischen der gemessenen Einstellung und anderen Konstrukten vermuten lassen. Um die "construct validity" einer Einstellungsskala zu demonstrieren, muß nun aufgezeigt werden, daß diese Beziehungen tatsächlich v o r l i e g e n . Wenn das nicht der Fall ist, so steht man vor dem bereits in Abschnitt 23 geschilderten Dilemma:¹⁾ Man kann sich dafür entscheiden, das operational-definierte Konstrukt K' nicht als Äquivalent des entsprechenden konstitutiv definierten Konstruktes K anzuerkennen, oder man kann die Hypothese über die Beziehung zwischen den konstitutiv definierten Konstrukten als falsifiziert ansehen. Angesichts dieses Dilemmas wird man im Regelfall zu einer Modifizierung sowohl des Meßinstrumentes als auch der Hypothesen über die Beziehungen zwischen den Konstrukten kommen. Bei einer Modifizierung des theoretischen Konstruktes kann man zu neuen Annahmen über die Beziehungen zwischen einer bestimmten Position auf der zu prüfenden Einstellungsskala und anderen Konstrukten und ebenfalls zu neuen Vorstellungen von den nötigen Meßoperationen gelangen: "The determination of construct validity, then, is not so much a process as it is a program".²⁾

1) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 7 f.

2) GHISELLI, E. E.: Theory of Psychological Measurement, a.a.O., S. 349

Obwohl die Bestimmung der "construct validity" weitgehend aufgrund vorliegender Daten erfolgt, wird sie doch auch letztlich in Form eines subjektiven Urteils ausgedrückt: Der Validitätsgrad kann nicht wie bei der "criterion validity" durch einen einzigen Index (Validitätskoeffizient) wiedergegeben, sondern muß verbal ausgedrückt werden.¹⁾

Die Beschreibung der Vorgehensweise bei der "construct validity" zeigt, daß hier mit mindestens zwei unsicheren Annahmen gearbeitet wird: Zum einen müssen die Hypothesen bestätigt sein, die das Konstrukt, dessen Messung auf seine Validität geprüft werden soll, mit anderen Konstrukten verbinden. Zum anderen muß die valide Messung dieser anderen Konstrukte unterstellt werden. Die Berechtigung beider Annahmen kann jedoch nicht endgültig nachgewiesen werden. Auch bei der Prüfung auf "construct validity" läßt sich demnach nur Evidenz für die Validität eines Instrumentes sammeln. Das bedeutet, daß alle drei hier dargestellten Verfahren der Validitätsprüfung nur zu näherungsweisen Lösungen führen können.

Abschließend sei im Zusammenhang mit der Erörterung des Validitätsproblems noch kurz auf die Frage nach der Zahl der mit einer Einstellungsskala gemessenen Dimensionen eingegangen: Viele Bemühungen bei der Konstruktion von Skalen sind darauf gerichtet, eindimensionale Instrumente zu konstruieren. Wieweit eine Skala eindimensional ist, hängt weitgehend von der Technik ab, die bei ihrer Konstruktion

1) Vgl. GHISELLI, E. E.: Theory of Psychological Measurement, a.a.O., S. 350

angewendet wird.¹⁾ Die bisherigen Versuche auf diesem Gebiet²⁾ zeigen, daß perfekte Eindimensionalität nicht erreicht werden konnte. Es handelt sich hierbei insofern um ein Validitätsproblem, als im Falle nicht perfekter Eindimensionalität ein Störfaktor oder mehrere Störfaktoren zu einer systematischen Verzerrung der Meßergebnisse führen. Die Bemühungen um Eindimensionalität bei der Skalenkonstruktion dienen also der Erhöhung der Validität.

Die bewußte Konstruktion mehrdimensionaler Skalen erscheint nur dann sinnvoll, wenn - z.B. im Falle von Einstellungsskalen - die Beiträge der verschiedenen Faktoren zu der bei einer Messung erreichten Gesamtposition in einem mehrdimensionalen Raum getrennt erfaßt werden können. Der Prozeß der Zuordnung von Zahlen ist im Falle ein- und mehrdimensionaler Skalierung grundsätzlich derselbe.³⁾ Die in der vorliegenden Arbeit untersuchten Skalierungstechniken werden lediglich daraufhin geprüft, wieweit sie in der Lage sind, Richtung und Intensität von Einstellungen auf dem Niveau von Ordinalskalen zu messen. Die Diskussion der speziellen Schwierigkeiten, die bei der Einbeziehung weiterer Dimensionen auftreten können, muß einer späteren Untersuchung überlassen bleiben.

333 Abgeleitete Messung (derived measurement)

Bei der abgeleiteten Messung⁴⁾ erlangen die Charakteristika Ordnung, Abstand und Ursprung Bedeutung

1) Vgl. GREEN, B. F.: Attitude Measurement, in: G. LINDZEY (Hrsg.): Handbook of Social Psychology, 1. Bd., Reading, Mass. und London 1954, S. 363

2) Vgl. z.B. die entsprechenden Ausführungen im Abschnitt 433 dieser Arbeit

3) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 37 - 40 und S. 247 ff.

4) Vgl. zum folgenden vor allem dens., a.a.O., S. 21 ff.

durch Gesetze ("laws") oder definitorische Setzungen ("stipulations"), die die zu messende Eigenschaft eines K-Konstruktes mit anderen K-Konstrukten in Beziehung setzen: "By derived measurement we understand the determination of a metrical scale by means of criteria which presuppose at least one previous scale of measurement."¹⁾ So gemessen, besitzen die Konstrukte direkte konstitutive Bedeutung, erhalten operationale Bedeutung jedoch nur indirekt über die Messung anderer Konstrukte. Die abgeleitete Messung eines Konstruktes setzt demnach die valide, grundlegende Messung zumindest eines anderen Konstruktes voraus. Es wurde bereits darauf hingewiesen, daß die Erfassung der Konstrukte K_5 und K_6 der schematischen Darstellung im Abschnitt 23 nur auf diese Weise möglich ist.²⁾ Das klassische Beispiel einer abgeleiteten Messung in der Physik ist die Messung der Dichte: Der Wert des Verhältnisses von Masse zu Volumen einer Substanz wird als Maß der Dichte verwendet.

Mit zunehmender Entwicklung einer Wissenschaft wird mehr und mehr von der Form der abgeleiteten Messung Gebrauch gemacht. Da aber in den Sozialwissenschaften noch keine Theorie besteht, die es ermöglicht, neue Konstrukte auf dem Wege über die abgeleitete Messung einzuführen, ist man vorläufig noch auf die Formen der Messung per definitionem

1) HEKPEL, C. G.: Fundamentals of Concept Formation in empirical Science, a.a.O., S. 69 f.
HEKPEL bezieht seine Definition zwar nur auf Skalen mit metrischen Eigenschaften; seine Definition kann jedoch auch für andere Meßniveaus gelten.

2) Vgl. Figur 1, S. 17

und die grundlegende Messung angewiesen, wobei jede spezielle Skala durch eine Mischung verschiedener Arten der Messung entstehen kann.¹⁾

1) Erwähnt sei noch, daß SUPPES und ZINNES noch eine dritte Art der Messung diskutierten, die sie "pointer measurement" nennen. Sie verstehen darunter "a numerical assignment (either fundamental or derived) based on the direct readings of some validated instrument." Vgl. SUPPES, P. und J. L. ZINNES: Basic Measurement Theory, a.a.O., S. 20 ff. Der Fall der "pointer measurement" wird u.a. am Beispiel der Waage erläutert, bei der man am Ausschlag des Zeigers direkt das Gewicht eines in der Waagschale befindlichen Gegenstandes ablesen kann. Eine solche Art der Messung dürfte jedoch bei den hier erörterten Konstrukten bis auf weiteres unmöglich sein.

4 Techniken der Skalenkonstruktion

4.1 Vorbemerkungen

Items¹⁾ sind die Elemente der in dieser Arbeit diskutierten Einstellungsskalen. GREEN definiert den Terminus folgendermaßen: "The particular situation or object that evokes the response, together with a specified set of response categories, is called an *i t e m*."²⁾ Die Zahl, die den Versuchspersonen bei der Anwendung einer Skala aufgrund ihrer Stellungnahmen zu den Items zuerteilt wird, nennt man Skalen- oder Punktwert. Ebenso wird die Zahl benannt, die der Position eines Items auf einem Einstellungskontinuum zugeordnet wird. Zwei Fragen sind im Zusammenhang mit den bei der Skalenkonstruktion verwendeten Items von besonderer Bedeutung: Erstens die Frage nach dem Inhalt der Items und zweitens die Frage nach der Art, wie ein Item mit der zu messenden Variablen verbunden ist. GREEN spricht hier von "item type" und "item form". Auf beide Fragen soll an dieser Stelle kurz eingegangen werden.

Die Formulierung der Items ist der Schritt bei der Konstruktion einer Skala, der am meisten von den Erfahrungen und der Intuition des jeweiligen Forschers abhängt. Es gibt zwar Regeln, die bei der Formulierung zu beachten sind,³⁾ sie sichern jedoch keineswegs den Erfolg. Die am meisten verwendeten Items enthalten Behauptungen über bestimmte Ein-

1) Der englische Ausdruck "Item" ist kaum zu übersetzen; er bedeutet etwa "Einzelelement" oder "Einzelfrage". Vgl. das Glossar in KÜNIG, R. (Hrsg.): Das Interview, Praktische Sozialforschung I, Köln und Berlin, 4. Aufl., 1965, S. 362.

Da der Terminus jedoch inzwischen auch im Deutschen gebräuchlich geworden ist, wird er in dieser Arbeit beibehalten.

2) GREEN, B. F.: Attitude Measurement, a.a.O., S. 338. Zu den folgenden Ausführungen vgl. ebenfalls GREEN, B. F.: Attitude Measurement, a.a.O., S. 338 und 341 ff.

3) Vgl. PAYNE, S. L.: The Art of Asking Questions, Princeton, N.J., 1951
Vgl. EDWARDS, A. L.: Techniques of Attitude Scale Construction, New York 1957, S. 13 f.

stellungsobjekte, sie sind "opinion statements". Solche Behauptungen können etwa lauten: "Alles, was in den Zeitungen steht, ist gelogen!", "Die Neger stinken!" oder "Das Fernsehen sollte die Sendezeit für politische Informationssendungen verdoppeln!" Die besondere Problematik dieser Behauptungen bei der Konstruktion von Einstellungsskalen besteht darin, daß die Bedeutung der verwendeten Wörter für verschiedene Gruppen von Versuchspersonen unterschiedlich sein kann. Deswegen ist es z.B. angebracht, bei der Konstruktion einer Skala die Probanden aus dem Kreis von Versuchspersonen auszuwählen, für den die Skala konstruiert werden soll.¹⁾ Ebenso kann sich die Bedeutung der verwendeten Wörter im Laufe der Zeit ändern. Aus diesem Grunde ist es bei der Verwendung älterer Skalen nötig zu prüfen, ob die diesbezüglichen Voraussetzungen noch in der Weise wie zum Zeitpunkt der Skalenkonstruktion gegeben sind.

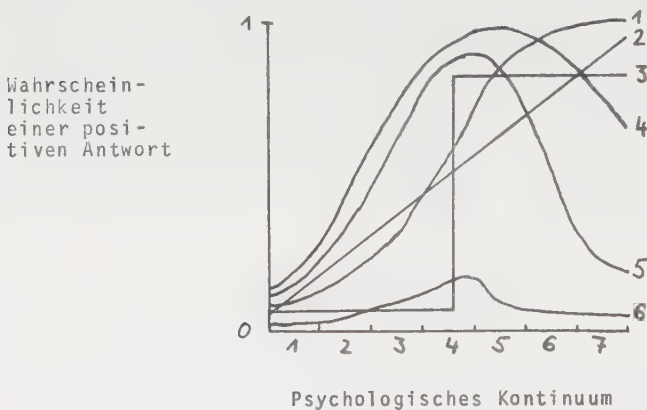
Die Reaktionen der Versuchspersonen hängen natürlich sehr stark vom Wortlaut der Items ab: Änderungen im Wortlaut können zu Änderungen in den Reaktionen führen. Deswegen ist es von besonderer Bedeutung, daß jede Versuchsperson mit genau demselben Wortlaut konfrontiert wird. Ein besonderes Problem besteht darin, daß viele Personen bereit sind, zu Behauptungen Stellung zu nehmen, die an sich für sie sinnlos sind. Bei der Formulierung von Items ist daher darauf zu achten, daß diese sowohl für die potentiellen Versuchspersonen als auch für die zu messende Einstellung verständlich sind.²⁾

Diese eher praktischen Erwägungen sind für die Diskussion der Skalierungstechniken weniger wichtig als die Frage, wie ein Item mit der zu messenden Variablen verbunden ist:

1) Vgl. die Ausführungen in Abschnitt 432

2) Vgl. GREEN, B. F.: Attitude Measurement, a.a.O., S. 341

Jedes Skalierungsmodell macht bestimmte Annahmen über die Art der Verbindungen zwischen den Items und den zu erfassenden Variablen; diese Verbindungen können mit Hilfe von Reaktionskurven dargestellt werden:¹⁾ Diese Kurven zeigen die empirisch zu ermittelnde Wahrscheinlichkeit einer positiven Reaktion auf ein Item als Funktion der Position auf dem latenten Einstellungskontinuum. Die beiden grundlegenden Formen von Reaktionskurven lassen sich anhand der folgenden graphischen Darstellung verdeutlichen:



Figur 2: Beispiele für Items mit monotonen und nichtmonotonen Reaktionskurven:
1, 2 und 3 sind monotone,
4, 5 und 6 nichtmonotone Reaktionskurven

Man kann zwischen Items mit m o n o t o n e n und n i c h t m o n o t o n e n Reaktionskurven²⁾ unter-

1) Vgl. GREEN, B. F.: Attitude Measurement, a.a.O., S. 343 f. GREEN spricht von "item operating characteristic", LAZARSFELD von "trace line"; vgl. hierzu den Abschnitt 437 dieser Arbeit.

2) Synonym hierzu werden in der Literatur die Ausdrücke "kumulative" bzw. "nichtkumulative" Items verwendet.

scheiden: Monotone Items sind dadurch charakterisiert, daß die Wahrscheinlichkeit einer positiven Reaktion als Funktion der Position, die eine Versuchsperson auf dem psychologischen Kontinuum einnimmt, ansteigt oder gleichbleibt, dagegen niemals sinkt. Ein Beispiel hierfür ist das Item: "Die Geburtenkontrolle ist besonders für diejenigen Frauen wünschenswert, die ihren Lebensunterhalt selbst verdienen müssen."¹⁾ Dieses Item ist mit fünf Antwortkategorien von "völlige Übereinstimmung" bis "völlige Ablehnung" versehen. Wenn es eine monotone Reaktionskurve aufweist, so bedeutet das, daß eine Versuchsperson auf dieses Item um so wahrscheinlicher positiv reagiert, je negativer ihre Einstellung gegenüber der Geburtenkontrolle ist. Diese Reaktionskurve wird in etwa durch die Kurve 1 in der Figur 2 wiedergegeben, wenn das psychologische Kontinuum von links nach rechts mit den Punkten 1 bis 7 eine immer positivere Einstellung zur Geburtenkontrolle darstellt. Bei einem nichtmonotonen Item dagegen verläuft die Reaktionskurve zunächst bis zu einem bestimmten Wert auf dem psychologischen Kontinuum steigend, um danach, wenn der Wert auf dem Kontinuum noch weiter steigt, zu fallen. Beispielhaft hierfür kann folgendes Item sein: "Ich gehe gerne ab und zu ins Kino, aber die gezeigten Filme enttäuschen mich manchmal."²⁾ Hierbei handelt es

1) Das Item stammt aus einer Skala, die die Einstellung gegenüber der Geburtenkontrolle messen soll; sie wurde mit Hilfe der Technik der summierten Einschätzungen konstruiert.

Vgl. WILKE, W. H.: An Experimental Comparison of the Speech, the Radio, and the Printed Page as Propaganda Devices. Arch. Psychol. 169, 1934, abgedruckt bei SHAW, M. E. und J. M. WRIGHT, Scales for the Measurement of Attitudes, a.a.O., S. 136 f.

2) Dieses Item wurde einer von THURSTONE entwickelten Skala zur Messung der Einstellung zum Film entnommen. Vgl. THURSTONE, L. L.: A Scale for Measuring Attitude toward the Movies, J. Educ. Res. 22, 1930, S. 89 - 94, abgedruckt in SHAW, M. W. und J. M. WRIGHT: Scales for the Measurement of Attitudes, a.a.O., S. 140 ff.

sich um ein dichotomisches Item; das heißt, daß für die Versuchspersonen nur zwei Reaktionsmöglichkeiten, eine positive und eine negative, vorgegeben werden. Hat dieses Item eine nichtmonotone Reaktionskurve, so werden nur die Versuchspersonen positiv reagieren, deren Einstellung in dem Item genau oder annähernd zum Ausdruck kommt. Alle Versuchspersonen dagegen, die eine entweder positivere oder negativere Einstellung zum Kino haben, werden negativ reagieren. Die Kurve 5 der Figur 2 stellt eine entsprechende Reaktionskurve dar. - Auf die Unterscheidung zwischen monotonen und nichtmonotonen Items wird im folgenden wiederholt zurückgegriffen.

42 Einteilung der Skalierungstechniken

Es gibt eine Reihe von Kriterien, die sich zur Klassifizierung der zu erörternden Skalierungstechniken heranziehen lassen. So kann z.B. das mit einer Technik zu erreichende Meßniveau ebenso wie die Art der Messung zur Klassifizierung benutzt werden. Auch ist es möglich, Verfahren, die vergleichende Urteile verlangen, von solchen zu unterscheiden, bei denen von kategorialen Urteilen ausgegangen wird. Im Rahmen dieser Arbeit soll lediglich der Klassifikationsversuch TORGERSONs¹⁾ erläutert werden, weil er als der theoretisch bedeutsamste erscheint. Die allgemeinen Ausführungen dieses Abschnitts werden durch die nachfolgende Diskussion der einzelnen Skalierungstechniken verdeutlicht.

TORGERSON geht bei seiner Klassifikation der Skalierungstechniken von den theoretischen Problemen aus, die dem jeweiligen Skalierungsversuch zugrunde liegen: "The

1) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 45 ff. et passim. Bei den in dieser Arbeit behandelten Skalierungstechniken werden als Stimuli Items verwendet.

distinction is based primarily on differences in the allocation of the variability of the responses to the stimuli."¹⁾ TORGERSON gelangt zu seiner Klassifikation, indem er die Frage stellt: Sind die Variationen in den Reaktionen, welche die Versuchspersonen auf die ihnen vorgelegten Items zeigen, auf die Versuchspersonen selbst, auf die Items oder auf Versuchspersonen und Items zugleich zurückzuführen? Auf diese Weise erhält er folgende Dreiteilung:

- a) Personen-orientiertes Messen ("subject-centered approach"),
 - b) Indikator-orientiertes Messen ("stimulus-centered" oder "judgment approach") und
 - c) Reaktions-orientiertes Messen ("response approach").²⁾
- Diese Dreiteilung beruht also auf einer unterschiedlichen Interpretation der systematischen Variation in den Reaktionen der Versuchspersonen auf ihnen vorgelegte Items.

421 Personen-orientiertes Messen

Die systematische Variation in den Reaktionen der Versuchspersonen auf die Items wird bei diesem Skalierungsansatz individuellen Unterschieden zwischen den Versuchspersonen zugeschrieben.³⁾ Anders ausgedrückt "werden die Unterschiede in den Werten als Differenzen zwischen den Versuchspersonen interpretiert"⁴⁾. Nur den Versuchspersonen

1) TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 46

2) Diese Übersetzung und Schreibweise der Ausdrücke TORGERSONs befindet sich bei SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, a.a.O., S. 351

3) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 46

4) SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, a.a.O., S. 351

sonen werden bei diesem Ansatz Skalenwerte zugeordnet, die Items selbst dagegen werden nicht skaliert. Sie sind als Replikationen anzusehen. Durch Hinzufügen oder Weglassen von Items aus demselben "Bedeutungskollektiv"¹⁾ nach dem Zufallsprinzip werden keine anderen Wirkungen auf die Ergebnisse der Skalierung ausgeübt als diejenigen, die auf die üblichen Samplefluktuationen zurückführbar sind.²⁾ Zwar können den einzelnen Items unterschiedliche Gewichtungen zuerteilt werden; das ist jedoch nur möglich aufgrund ihrer Fähigkeit, zwischen den Versuchspersonen zu diskriminieren und nicht auf der Grundlage ihrer relativen Position auf einem Einstellungskontinuum.³⁾

Wie TORGERSON bemerkt, hat diese Vorgehensweise nicht zu irgendwelchen bedeutenden Skalierungsmodellen geführt; meist handelt es sich hierbei um Messung per definitionem. Die meisten Anwendungsbeispiele für solche Verfahren finden sich auf dem Gebiet des sogenannten "mental testing".⁴⁾ Das einzige auf diesem Ansatz aufbauende Verfahren, das sich für die Messung bestimmter im Massenkommunikationsprozeß relevanter Einstellungen von Kommunikatoren und Rezipienten eignet, ist die von LIKERT entwickelte Technik der summierten Einschätzungen.⁵⁾ Bezeichnenderweise hat LIKERT seiner Technik kein eigentliches Skalierungsmodell zugrunde gelegt. Spätere Versuche, seine Technik als auf dem sogenannten "single common factor model" beruhend zu interpretieren, sind nachträgliche Konstruktionen von mehr oder weniger großer Plausibilität.

1) Zu diesem Terminus vgl. die Ausführungen in Kapitel 433 dieser Arbeit.

2) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 46

3) Vgl. ders., a.a.O., S. 47

4) Vgl. z.B. CROWBACH, L. J.: Essentials of Psychological Testing, a.a.O.
GULLIKSEN, H.: Theory of Mental Tests, New York, London und Sidney 1950

5) Vgl. hierzu das Kapitel 432 dieser Arbeit.

422 Indikator-orientiertes Messen:

Bei dieser Vorgehensweise wird die systematische Variation in den Reaktionen der Versuchspersonen auf ihnen vorgelegte Items den Unterschieden zwischen den Items in bezug auf die zu messende Einstellung zugeschrieben.¹⁾ Das bedeutet, daß man "die Differenzen in den Beobachtungswerten auf Unterschiede in der Bedeutung der Indikatoren"²⁾ zurückführt. In diesem Falle ist zunächst die Skalierung der Items beabsichtigt, wobei die Versuchspersonen als Replikationen angesehen werden: Durch ein Hinzufügen oder Weglassen von Versuchspersonen aus einer möglichen Gesamtpopulation nach dem Zufallsprinzip entsteht keine andere Wirkung auf Ergebnisse der Skalierung als diejenige, die auf die üblichen Samplefluktuationen zurückgeführt werden kann.³⁾ Nur den Items werden also bei der Konstruktion der Skala Werte zuerteilt.

Beispiele für diese Vorgehensweise sind die in Abschnitt 431 erläuterten "Judgment"-Techniken. Die von den Versuchspersonen zu erfüllende Aufgabe besteht in einem solchen Fall darin, ihnen vorgelegte Items hinsichtlich einer zu messenden Einstellung zu bewerten: "The subject responds to a stimulus with respect to its relation among other stimuli in the defined continuum."⁴⁾ Weil die Versuchspersonen die Items im Hinblick auf deren relative Position zu anderen Items beurteilen, sollen die Wirkungen einer eventuellen Voreingenommenheit der Versuchspersonen selbst minimal sein:

-
- 1) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 46
 - 2) SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, a.a.O., S. 352
 - 3) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 46
 - 4) ders., a.a.O., S. 48.
COOMBS spricht hier in seiner Terminologie von "Aufgabe A". Vgl. z.B. COOMBS, C. H.: Theory and Methods of Social Measurement, a.a.O., S. 491 ff.

"Since the subject responds to the stimulus on the basis of its relativ position among other stimuli, the effect of his own bias is minimized."¹⁾ Auf die Problematik dieser Annahme wird vor allem im Zusammenhang mit der Erörterung der Technik der gleich~~er~~scheinenden Abstände in Abschnitt 4312 näher eingegangen.

Beim Indikator-orientierten Messen bestehen nach TORGERSON kaum grundlegende theoretische Schwierigkeiten. Praktische Schwierigkeiten für das Ordnen der Items können in der Versuchssituation dagegen vor allem dann auftreten, wenn die Items auf dem psychologischen Kontinuum eng beeinander liegen, oder wenn die Einstellung, auf die sich die Versuchspersonen bei ihrer Beurteilung beziehen sollen, nicht klar definiert ist. Besonders schwierig wird diese Situation dann, wenn der Versuchsleiter - was in der Regel der Fall sein dürfte - nicht in der Lage ist, wiederholte unabhängige Urteile über jedes Item einzuholen.²⁾

Ein wesentlicher Unterschied zwischen dem Indikator-orientierten Messen und dem bereits erläuterten Personen-orientierten Messen besteht darin, daß beim Personen-orientierten Messen zumeist Messung per definitionem vorliegt, während beim Indikator-orientierten Messen zumeist eine fundamentale Messung versucht wird. Dabei ist in der Regel die Konstruktion einer Intervallskala beabsichtigt; die Frage nach Ordnung und Abstand der Items einer Skala wird zum Gegenstand empirischer Prüfung.

Der Unterschied in der Vorgehensweise kann, was die jeweilige Versuchsanordnung betrifft, an folgendem Beispiel kurz erläutert werden. Zwei Items, die Einstellungen gegenüber dem Film ausdrücken, sollen zu diesem Zweck heran-

1) TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 48

2) Vgl. TORGERSON, a.a.O., S. 52 f.

gezogen werden¹⁾):

a) "Der Film ist die wichtigste Ursache des Verbrechens."

b) "Der Film ist heutzutage die lebendigste Form der Kunst."

Im Falle des Personen-orientierten Messens werden die Items den Versuchspersonen mit der Anweisung vorgelegt, aufgrund ihrer eigenen Einstellung zum Film anhand eines vorgegebenen Antwortschemas zu den Items Stellung zu nehmen; die Versuchspersonen sollen also ihre Ablehnung oder Zustimmung zu den in den Items ausgedrückten Einstellungen äußern. Die Antworten der Versuchspersonen werden dann lediglich in der Absicht analysiert, diejenigen Items herauszufinden, die am stärksten zwischen Personen mit positiver und Personen mit negativer Einstellung zum Film diskriminieren. Beim Indikator-orientierten Messen dagegen besteht die Aufgabe der Versuchspersonen darin, die Positionen der ihnen vorgelegten Items auf einem psychologischen Kontinuum ohne Berücksichtigung der jeweils eigenen Einstellung zu dem betreffenden Einstellungsobjekt (Film) anzugeben. Die Antworten der Versuchspersonen werden diesmal mit der Absicht analysiert, den Items Skalenwerte zuzuordnen, die ihre Position auf dem psychologischen Kontinuum wiedergeben.²⁾ Die Eigenschaften der Skala (Ordnung und Abstand der zuerteilten Skalenwerte) werden in diesem Falle empirisch ermittelt.

Bei dem Versuch, mittels Indikator-orientierter Messung eine Intervallskala zu konstruieren, lassen sich zwei

1) Die Items sind in der bereits erwähnten Skala von THURSTONE enthalten: Vgl. THURSTONE, L. L.: A Scale for Measuring Attitude toward the Movies, abgedruckt in: SHAW, H. E. und J. M. WRIGHT: Scales for the Measurement of Attitudes, a.a.O., S. 140 - 142

2) Bei einem vorgegebenen Kontinuum von sechs Punkten zur Einstufung der Items wurde im vorliegenden Fall für das Item a ein Skalenwert von 0,2 und für das Item b ein Skalenwert von 4,3 errechnet, wobei 0 die negativste und 5 die positivste Einstellung gegenüber dem Film ausdrücken soll.

verschiedene Vorgehensweisen unterscheiden:

- a) Wenn die Intervalle aufgrund direkter quantitativer Urteile über die Items im Hinblick auf die zu messende Einstellung gebildet werden, spricht TORGERSON vom "quantitative judgment approach".¹⁾ Die Anweisung für die Versuchspersonen kann in diesem Falle etwa lauten: "Ordnen Sie die Items den hier vorliegenden elf (mit A bis K benannten) Kategorien zu, wobei die Abstände zwischen den Kategorien gleich sein sollen!" Hiermit wird eine Anweisung erteilt, wie sie vor allem bei der Technik der gleich erscheinenden Abstände angewendet wird.²⁾ Sie bedeutet, daß die Versuchsperson nicht nur Urteile der Form "i ist größer als j" abgeben muß, wie das etwa bei der Technik des Paarvergleichs der Fall ist, sondern daß sie auch in der Lage sein muß, die psychologische Distanz oder das Verhältnis zwischen den einzelnen Items zu beurteilen: "It is assumed that he can rate, sort, or arrange members of the stimulus series in such a manner that the ratios of differences between the numbers assigned to the stimuli are equal to the ratios of the distance separating the stimuli on the psychological continuum."³⁾

1) Vgl. hierzu und zu den folgenden Ausführungen: TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 54 ff. Die englischen Ausdrücke für die verschiedenen Vorgehensweisen werden hier nicht übersetzt, da keine adäquaten deutschen Termini zur Verfügung stehen.

2) Diese Anweisung ist typisch für einen Unterfall des "quantitative judgment approach", den TORGERSON die "subjective-estimate" Methode nennt. Vgl. die Diskussion in Abschnitt 4312 dieser Arbeit. Die beiden weiteren Unterfälle des "quantitative judgment approach", die "fractionation"-Methode und "equissection"-Methode, werden nicht weiter erörtert, weil im folgenden keine entsprechenden Skalierungstechniken behandelt werden. Vgl. hierzu die zusammenfassende Darstellung bei TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., Kap. 5 und 6.

3) Torson, W. S.: S. 54.

- b) Bei der weiteren, innerhalb des Indikator-orientierten Messens möglichen Vorgehensweise wird versucht, die Variation in den Urteilen der Versuchspersonen in bezug auf jedes Item oder jede Itemkombination zur Konstruktion einer Intervallskala zu benutzen; TORGERSON nennt das den "variability judgment approach".¹⁾ Hierzu gehört z.B. die Technik des Paarvergleichs. Die Frage an die Versuchspersonen kann dann etwa lauten: "In welchem der beiden Items, i oder j, kommt die Einstellung x (z.B. Konservatismus) stärker zum Ausdruck?" Die Anwendung des "variability judgment approach" kann nur dann zur Bildung einer Intervallskala führen, wenn - von Urteilen über die bloße Ordnung der Items ausgehend - Hypothesen akzeptiert werden können, die eine Verbindung zwischen der Variabilität der Urteile über die einzelnen Items und der relativen Position dieser Items auf einem psychologischen Kontinuum herstellen. Ohne Variabilität in den Urteilen der Versuchspersonen kann mit Hilfe dieses Verfahrens von vornherein nur ein Instrument mit den Eigenschaften einer Ordinalskala konstruiert werden. Eine mögliche Hypothese bezieht sich beispielsweise darauf, daß die Skalenwerte der Items normal auf dem psychologischen Kontinuum verteilt sind. Die von THURSTONE entwickelten Techniken des Paarvergleichs und der nachträglich bestimmten Abstände²⁾, die Varianten dieses Ansatzes darstellen, zeichnen sich vor anderen beim "variability judgment approach" ebenfalls möglichen Techniken dadurch aus, daß das der Skalierung zugrunde gelegte Modell und damit auch die Eigenschaft gleicher Intervalle Gegenstand empirischer Prüfung werden: "If the

1) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 54, 55 und 56 - 58

2) Vgl. hierzu die Abschnitte 4311 und 4313 dieser Arbeit.

subjects do not behave in the manner required by the model, then the theory will not fit the experimental data, and the model will be rejected."¹⁾ Die Konstruktion einer Skala mit gleichen Intervallen gelingt nur dann, wenn die empirischen Daten mit dem Modell und den daraus abgeleiteten Gleichungen übereinstimmen.

423 Reaktions-orientiertes Messen

Bei dem dritten zu erörternden Skalierungsansatz wird die systematische Variation in den Reaktionen der Versuchspersonen auf die Items einer Skala sowohl den individuellen Unterschieden zwischen den Versuchspersonen als auch den Unterschieden zwischen den Items in bezug auf die zu messende Einstellung zugeschrieben.²⁾ Die Beobachtungswerte werden somit "als Information über Indikatoren und Versuchspersonen zugleich gedeutet"³⁾. Bei diesem Verfahren kann eine Skalierung sowohl der Items als auch der Versuchspersonen erreicht werden; in den meisten Fällen wird lediglich Ordinalskalenniveau angestrebt. Beispielhaft für diesen Ansatz ist die von GUTTMAN entwickelte Technik der Skalogramm-Analyse, wie sie in Abschnitt 433 diskutiert wird. Die Aufgabe der Versuchspersonen besteht in diesem Falle darin, "to respond to a stimulus on the basis of the position of the stimulus in relation to the subject's own position with respect to the attribute".⁴⁾ Wenn die Versuchspersonen also etwa die Aufgabe haben, aus

1) TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 58

2) Vgl. ders., a.a.O., S. 46

3) SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, a.a.O., S. 352

4) TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 48.

COOMBS bezeichnet in seiner Terminologie eine entsprechende Anweisung an die Versuchspersonen als "Aufgabe B". Vgl. COOMBS, C. H.: Theory and Methods of Social Measurement, a.a.O., S. 491 ff.

einer Reihe von Items dasjenige auszuwählen, das ihrer eigenen Einstellung am nächsten kommt, so wird diese Auswahl sowohl durch die Einstellung der betreffenden Versuchspersonen als auch durch die Positionen der Items auf dem psychologischen Kontinuum beeinflusst. Werden die Ausgangsdaten für die Skalierung auf die Weise gewonnen, so ist eine direkte Prüfung der Frage, ob die Postulate der Ordnung für eine gegebene Reihe von Items oder ein bestimmtes Sample von Versuchspersonen vorliegen, nicht möglich. Zur Prüfung dieser Frage ist vielmehr die Anwendung indirekter, abgeleiteter Verfahren nötig.¹⁾

Die verschiedenen Ansätze Reaktions-orientierten Messens unterscheiden sich vor allem in zweierlei Hinsicht:

- a) **D e t e r m i n i s t i s c h e** und **p r o b a b i l i s t i s c h e** Modelle unterscheiden sich in der Behandlung von Fehlern.²⁾ Im ersten Fall wird ein formales Modell errichtet, von dem gar nicht erwartet wird, daß es für die erhobenen Daten vollständig zutrifft. Für die Behandlung von abweichenden Fällen enthält das Modell keine Regelungen. Ein abweichender Fall in dem erhobenen Datenmaterial könnte demnach ein ausreichender Grund für die Zurückweisung des Modells sein. Da nun bei der Anwendung eines in der entsprechenden Weise konstruierten Meßinstrumentes praktisch immer abweichende Fälle auf-

1) Auf die Unterschiede in der Vorgehensweise bei der Herstellung der Ordnung der Items - und bei diesem Ansatz auch der Versuchspersonen - zwischen Indikator-orientiertem und Reaktions-orientiertem Messen soll an dieser Stelle nicht näher eingegangen werden. Vgl. hierzu die folgenden Abschnitte dieser Arbeit sowie zu den grundsätzlichen Überlegungen TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 49 - 52 und S. 58

2) Vgl. ders., a.a.O., S. 59

treten, würde das bedeuten, daß ein solches Modell immer zurückgewiesen werden müßte. Daher können diese Modelle nur als mehr oder weniger enge Annäherung an die Wirklichkeit angesehen werden. Es lassen sich Maßstäbe entwickeln, die etwas darüber aussagen, wieweit ein bestimmtes Modell für die erhobenen Daten adäquat ist; diese Maßstäbe müssen jedoch im Grunde willkürlich bleiben. Beispiele für deterministische Modelle Reaktions-orientierten Messens sind die Skalogramm-Analyse und die Technik der transferierten Einschätzungen.¹⁾ Bei der Skalogramm-Analyse dient z.B. der Reproduktionskoeffizient als Maß der Adäquanz des Modells für die erhobenen Daten, bei der Technik der transferierten Einschätzungen fehlen solche Maße noch weitgehend.

Bei den probabilistischen Modellen wird die Behandlung der Fehler bereits in das Modell einbezogen. In diesem Falle können daher eher statistische Maßstäbe entwickelt werden, die eine weniger willkürliche Entscheidung über Annahme oder Ablehnung eines Modells ermöglichen sollen.²⁾ Beispiel für ein probabilistisches Modell Reaktions-orientierten Messens ist die Analyse latenter Strukturen.³⁾

- 3) Die weitere Unterscheidung zwischen k a t e g o r i a - l e n und v e r g l e i c h e n d e n Urteilen beruht darauf, welche Aufgaben die Versuchspersonen jeweils zu erfüllen haben. Im Falle kategorialer Urteile müssen die Versuchspersonen über Zustimmung oder Ablehnung gegenüber einem Item entscheiden. Bei der zwei-

1) Vgl. die Abschnitte 433 und 435 der vorliegenden Arbeit.

2) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 59

3) Vgl. den Abschnitt 437 der vorliegenden Arbeit.

ten Art von Urteilen vergleichen die Versuchspersonen zunächst zwei oder mehr Items, um sie dann gegebenenfalls nach einem vorgegebenen Kriterium zu ordnen.¹⁾ Die Skalogramm-Analyse und die Analyse latenter Strukturen sind Beispiele für Reaktions-orientiertes Messen auf der Grundlage kategorialer Urteile, während die Versuchspersonen z.B. bei der Technik der transferierten Einschätzungen vergleichende Urteile abgeben müssen.

43 Erläuterung und Kritik der Techniken

431 Die "Judgment"-Techniken

Bei den "Judgment"-Techniken handelt es sich um Indikator-orientiertes Messen. Die Skalierungsaufgabe besteht darin, nichtmonotone Items auf einem psychologischen Kontinuum von positiv bis negativ zu plazieren. Allen drei hier dargestellten Verfahren ist gemeinsam, daß dazu die Urteile einer mehr oder weniger großen Gruppe von Versuchspersonen herangezogen werden, die als Experten oder "judges" fungieren. Die Techniken unterscheiden sich jedoch in der Art der von diesen Experten abgegebenen Urteile und in der Form, wie diese Urteile ausgewertet werden, um den Skalenwert eines Items auf einem psychologischen Kontinuum zu errechnen. Angestrebt wird die Konstruktion einer Intervallskala.

4311 Die Technik des Paarvergleichs

Die Technik des Paarvergleichs bildet die eigentliche Grundlage der verschiedenen "Judgment"-Techniken und soll

1) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 60.
Die Unterscheidung zwischen kategorialen und vergleichenden Urteilen läßt sich natürlich auch bei anderen Skalierungsansätzen, so beim Indikator-orientierten Messen, durchführen.

deshalb hier zuerst diskutiert werden. Diese Technik ist ebenso wie die folgenden mit dem Namen THURSTONE verbunden, der damit begonnen hat, den Paarvergleich für die Konstruktion von Skalen zur Messung von Einstellungen nutzbar zu machen.¹⁾ In der experimentellen Psychologie sind bereits vor den Arbeiten THURSTONEs Versuche unternommen worden, in denen die Beziehungen zwischen der Ordnung von Objekten (oder Stimuli) auf einem bekannten p h y s i s c h e n Kontinuum und der Ordnung derselben Objekte auf einem zunächst unbekannten p s y c h o l o - g i s c h e n Kontinuum - beruhend auf den Urteilen von Versuchspersonen - analysiert wurden. Bei diesen sogenannten psychophysischen Verfahren wurden z.B. Objekte gleicher Größe aber unterschiedlichen Gewichts gewogen und nach ihrem Gewicht geordnet (physisches Kontinuum). Dieselben Objekte wurden dann paarweise einer Anzahl von Versuchspersonen mit der Anweisung vorgelegt zu beurteilen, welches Objekt des betreffenden Paares schwerer sei. Auf der Grundlage der durchschnittlichen Urteile der Versuchspersonen ließ sich dann eine Ordnung der Objekte vom leichtesten bis zum schwersten auf einem psychologischen Kontinuum herstellen.

Die besondere Leistung THURSTONEs bei der Formulierung seines Gesetzes der vergleichenden Urteile besteht darin, daß er hiermit die Grundlagen dafür erarbeitete, eine Reihe von Stimuli auch dann entlang einem psychologischen Kontinuum zu ordnen, wenn kein bekanntes physisches Kontinuum besteht, zu dem die Werte der einzelnen Stimuli in Beziehung gesetzt werden könnten.²⁾ Bei der Entwicklung seines Modells für die hier zu beschreibenden "Judgment"-Techniken

1) Vgl. THURSTONE, L. L.: Psychophysical Analysis, Am. J. Psychol. 38, 1927, S. 368 - 389.
Vgl. ders., A Law of Comparative Judgment, Psychol. Rev. 34, 1927, S. 273 - 286.
Beide Aufsätze sind abgedruckt in dem Sammelband:
THURSTONE, L. L.: The Measurement of Values, a.a.O., S. 19 - 49.

2) Vgl. EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 20

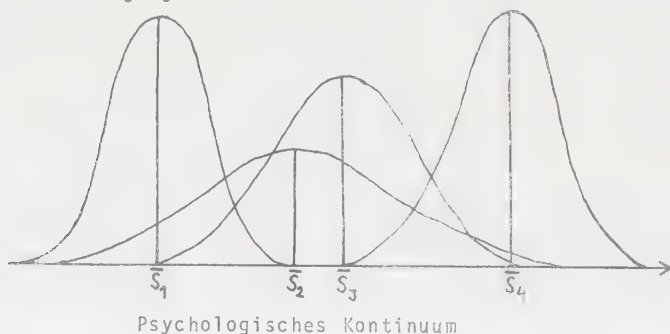
ging THURSTONE von folgenden Überlegungen aus: Ein psychologisches Kontinuum kann als Kontinuum subjektiver bzw. psychologischer Größen angesehen werden. Jede psychologische Größe entsteht aufgrund eines "Diskriminationsprozesses" ("discriminal process"). THURSTONE beschreibt diesen Diskriminationsprozeß als den Prozeß "by which the organism identifies, distinguishes, discriminates, or reacts to stimuli".¹⁾ TORGERSON bemerkt dazu, daß in dieser Beschreibung keine Annahmen über die Art, wie solche Prozesse - etwa physiologisch - ablaufen, enthalten sind. Die Einführung des Terminus "Diskriminationsprozeß" heiße lediglich, daß der vieldeutige Ausdruck "sensation" verwendet werde.²⁾

Es wird nun angenommen, daß jeder Stimulus i , wenn er einer Versuchsperson vorgelegt wird, einen Diskriminationsprozeß S_i hervorruft. Dabei ist es möglich, daß ein bestimmter Stimulus i nicht immer denselben Diskriminationsprozeß S_i hervorruft, sondern verschiedene, die unterschiedliche Werte auf dem psychologischen Kontinuum einnehmen können: So kann mit einem Stimulus i eine Reihe von Diskriminationsprozessen verbunden sein. Wenn nun dieser Stimulus einer Versuchsperson oftmals (oder vielen Versuchspersonen einmal) vorgelegt wird, so entsteht auf dem psychologischen Kontinuum eine Häufigkeitsverteilung von Diskriminationsprozessen, die mit diesem Stimulus i verbunden sind. Zusätzlich wird die Annahme eingeführt, daß

-
- 1) THURSTONE, L. L.: Psychophysical Analysis, in:
L. L. THURSTONE: The Measurement of Values, a.a.O.,
S. 20
 - 2) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling,
a.a.O., S. 156
Die Darstellung folgt hier im wesentlichen TORGERSONS
zusammenfassenden Ausführungen über THURSTONE'S Skalierungsmodell.

die Häufigkeiten der mit den Stimuli verbundenen Diskriminationsprozesse eine Normalverteilung auf dem psychologischen Kontinuum bilden.

Eine solche Normalverteilung läßt sich vor allem mit Hilfe zweier Maße beschreiben: einem Maß für den Mittelwert und einem Maß für die Streuung. Der Diskriminationsprozeß, der am häufigsten mit einem Stimulus i verbunden ist, wird modaler Diskriminationsprozeß ("modal discriminial process") $\bar{S}_i$ genannt. Als Skalenwert eines Stimulus i auf einem psychologischen Kontinuum wird nun der Wert dieses modalen Diskriminationsprozesses festgelegt.¹⁾ Die Standardabweichung der mit einem Stimulus i verbundenen Diskriminationsprozesse wird von THURSTONE als Diskriminationsstreuung ("discriminal dispersion") σ_i bezeichnet. Diese Diskriminationsstreuungen brauchen nicht für alle Stimuli gleich groß zu sein.²⁾ Die folgende Abbildung verdeutlicht die Grundüberlegungen THURSTONES:³⁾



Figur 3: Angenommene Normalverteilung von Diskriminationsprozessen, die von den Stimuli 1 bis 4 hervorgerufen werden

- 1) Weil bei einer Normalverteilung Modus, Median und arithmetisches Mittel zusammenfallen, kann als Skalenwert eines Stimulus auch der Median oder das arithmetische Mittel des entsprechenden Diskriminationsprozesses verwendet werden.
- 2) Vgl. THURSTONE, L. L.: Stimulus Dispersion in the Method of Constant Stimuli, J. Exp. Psychol. 15, 1932.

Die Hauptschwierigkeit bei der Konstruktion einer Skala und damit der Berechnung der Skalenwerte für die Items liegt darin, daß die Versuchspersonen, denen die Items vorgelegt werden, die Werte der Diskriminationsprozesse auf dem psychologischen Kontinuum nicht direkt angeben können. Die Berechnung der Skalenwerte muß daher indirekt erfolgen. Zu diesem Zweck können Hypothesen in Form von Gleichungen formuliert werden, die die Urteile über die Beziehungen zwischen Stimuli, die eine Versuchsperson abgeben kann, in ein Verhältnis zu den Skalenwerten und Streuungen von Stimuli auf einem psychologischen Kontinuum setzen. Anhand dieser Gleichungen können dann die Skalen- und die Streuungswerte für die zu skalierenden Stimuli errechnet werden.¹⁾ Eine Möglichkeit, solche Gleichungen zu bilden, zeigt das in diesem Kapitel zu behandelnde Gesetz der vergleichenden Urteile ("law of comparative judgments"), bei dem der Skalierung Urteile in Form des Paarvergleichs zugrundegelegt werden. Eine weitere Möglichkeit wird durch das Gesetz des kategorialen Urteils ("law of categorial judgment") bezeichnet, das eine Modifikation des Gesetzes der vergleichenden Urteile darstellt; von dieser Modifikation geht die Technik der nachträglich bestimmten Abstände aus.²⁾

S. 284 - 297.

Vgl. BURROS, R. H.: The Application of the Method of Paired Comparisons to the Study of Reaction Potential, Psychol. Rev. 58, 1951, S. 60 - 65.

- 3) Die Abbildung wurde in leicht geänderter Form übernommen aus: TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 157

1) Vgl. ders., a.a.O., S. 157 f.

2) Vgl. den Abschnitt 4313 dieser Arbeit.

Auf eine kurze Zusammenfassung gebracht, behauptet das Gesetz der vergleichenden Urteile, "daß die Werte für jeden Stimulus, wenn dieser wiederholt von einer einzelnen Person mit anderen Stimuli verglichen wird, eine Gaußkurve bilden, deren Scheitelpunkt (im folgenden bezeichnet mit $\bar{S}$ als Skalenwert des betreffenden Stimulus; Anm. d. Verf.) als Position auf einem psychologischen Kontinuum gedeutet werden kann. Ersetzt man die wiederholten Vergleiche bei einer Versuchsperson durch jeweils einmalige Vergleiche bei einer Vielzahl von Personen, so wird die gleiche und auf die gleiche Weise deutbare Gesetzmäßigkeit bei der Verteilung der Werte erwartet."¹⁾

Die bei den "Judgment"-Verfahren verwendeten Stimuli sind dichotomische Items, zumeist in Form von "opinion statements".²⁾ Wenn nun eine Reihe von n Items formuliert ist, so wird davon ausgegangen, daß jedes dieser Items in einem unbekannten Ausmaß die Einstellung mißt, die gemessen werden soll. Das Problem liegt also darin, den einzelnen Items Punktwerte so zuzuteilen, daß das Ausmaß, in dem sie die zu messende Einstellung auf dem psychologischen Kontinuum von positiv bis negativ repräsentieren, wiedergegeben wird. Im Rahmen dieses Abschnittes sollen lediglich die Grundlagen der Verfahren zur Errechnung der Skalenwerte für die verwendeten Items und die Möglichkeiten der Anwendung der mit Hilfe dieser Technik konstruierten

1) SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, a.a.O., S. 355.
SCHEUCH bemerkt an derselben Stelle, daß hier "eher eine definitorische Setzung für die Anwendbarkeit des Verfahrens als ein Gesetz" vorliege.

2) Vgl. GREEN, B. F.: Attitude Measurement, a.a.O., S. 344

Skalen erörtert werden. Auf eine Reihe von Einzelheiten und Modifikationen muß dabei - wie bereits in der Einleitung ausgeführt - ebenso verzichtet werden, wie auf die Beschreibung verschiedener Signifikanztests für das Verfahren des Paarvergleichs.¹⁾

Die Technik des Paarvergleichs kann vor allem dann angewendet werden, wenn nur relativ wenige Items skaliert werden sollen. Jedes Item wird mit jedem anderen Item kombiniert und den Versuchspersonen jeweils paarweise zur Stellungnahme vorgelegt. Das bedeutet, daß jede Versuchsperson $n(n-1)/2$ Urteile abgeben muß, wenn n die Zahl der Items ist. 20 Items erfordern somit bereits 190 Urteile, 40 Items 780 Urteile.²⁾ Bei der Berechnung der Skalenwerte für die einzelnen Items wird folgendermaßen vorgegangen:³⁾

- 1) Vgl. hierzu besonders die Arbeiten von MOSTELLER, F.: Remarks on the Method of Paired Comparisons:
 - I. The Least Squares Solution Assuming Equal Standard Deviations and Equal Correlations, Psychometrika 16 1951, S. 3 - 10;
 - II. The Effect of an Aberrant Standard Deviation when Equal Standard Deviations and Equal Correlations Are Assumed, Psychometrika 16, 1951, S. 203 - 206;
 - III. A Test of Significance for Paired Comparisons when Equal Correlations and Equal Standard Deviations Are Assumed, Psychometrika 16, 1951, S. 207 - 218Vgl. auch die zusammenfassende Darstellung bei EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 53 - 81
- 2) Um die Zahl der Urteile pro Versuchsperson zu senken, sind Formen des sogenannten unvollständigen Paarvergleichs entwickelt worden. Jede Versuchsperson muß dabei nur einen Teil der Itempaare beurteilen, wobei die Verteilung dieser Paare auf die Versuchspersonen nach dem Zufallsprinzip erfolgt.
Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 191 - 194
Zur Art der geforderten Urteile vgl. weiter unten.
- 3) Die Erläuterung der Berechnungsweise für die Skalenwerte der Items bei vollständigen und unvollständigen Datenmatrizes sowie einer Möglichkeit der Prüfung der Daten auf innere Konsistenz folgt weitgehend der Darstellung bei EDWARDS.
Vgl. EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 23 ff.

Wenn zwei Items i und j , die sich auf ein bestimmtes Einstellungsobjekt beziehen, formuliert sind, so werden diese Items einer Gruppe den Versuchspersonen vorgelegt, die vergleichende Urteile darüber abgibt, ob in Item i oder in Item j eine positivere Einstellung gegenüber dem Einstellungsobjekt ausgedrückt wird, wobei neutrale Urteile ausgeschlossen sind. Die Versuchspersonen sollen also nicht aufgrund ihrer eigenen Einstellung zu dem betreffenden Einstellungsobjekt urteilen. Wenn sich nun ergibt, daß 50 % der Versuchspersonen Item i positiver als Item j einschätzen und die anderen 50 % der Versuchspersonen Item j positiver als Item i beurteilen, so wird, um einen Punktwert für die Items berechnen zu können, daraus geschlossen, daß die Items i und j genau gleich positiv in bezug auf die zu messende Einstellung sind. Sie werden also denselben Skalenwert bekommen: $\bar{S}_i = \bar{S}_j$. Wenn mehr als 50 % der Versuchspersonen i positiver als j beurteilen, so ergibt sich aufgrund derselben Überlegung, daß Item i einen höheren Skalenwert als Item j erhält: $\bar{S}_i > \bar{S}_j$. Im umgekehrten Falle, wenn mehr als 50 % der Versuchspersonen j positiver als i einschätzen, wird Item j auf den höheren Skalenwert gesetzt: $\bar{S}_i < \bar{S}_j$. Die Zuerteilung der Skalenwerte für die einzelnen Items ist somit eine Funktion des Verhältnisses von vergleichenden Urteilen der Form $i > j$.

Wenn die Items i und j von den Versuchspersonen in der beschriebenen Weise beurteilt wurden, so ergibt sich eine empirische Häufigkeitsverteilung der Form:

$$(1) \quad f_{ij} = i > j$$

Dabei ist f_{ij} die Häufigkeit, mit der i positiver als j eingeschätzt worden ist. Die so gefundenen Häufigkeiten können zur Zahl der Gesamturteile ins Verhältnis gesetzt werden:

$$(2) \quad p_{ij} = \frac{f_{ij}}{N}$$

N ist dabei die Anzahl aller Urteile und p_{ij} der Anteil der Urteile $i > j$ an der Gesamtzahl aller Urteile. Die Werte von p_{ij} können nun mit Hilfe einer Normalverteilungstabelle in Z-Werte, z_{ij} , umgeformt werden. Wenn beispielsweise $p_{ij} = 0,842$ (das heißt 84,2 % aller vergleichenden Urteile lauten $i > j$), dann ist $z_{ij} = 1,003$; wenn $p_{ij} = 0,182$ (das heißt 18,2 % aller vergleichenden Urteile lauten $i > j$), dann ist $z_{ij} = -0,908$.¹⁾

Da davon ausgegangen wurde, daß die Werte für alle Stimuli (hier die Items i und j) bei wiederholten Beurteilungen eine Normalverteilung bilden, so ist die Differenz zwischen diesen Normalverteilungen ebenfalls normal verteilt und besitzt eine Standardabweichung von:

$$(3) \quad \sigma_{i-j} = \sqrt{\sigma_i^2 + \sigma_j^2 - 2r_{ij} \sigma_i \sigma_j}$$

Dabei ist: σ_{i-j} = Standardabweichung der Differenzen,
 $S_i - S_j$

σ_i = Standardabweichung von S_i

σ_j = Standardabweichung von S_j

r_{ij} = Korrelation zwischen S_i und S_j

THURSTONE berechnet nun die Trennung zwischen zwei Skalwerten $\bar{S}_i$ und $\bar{S}_j$ mit Hilfe der Formel (3) und dem z_{ij} -Wert, der dem p_{ij} -Wert für die vergleichenden Urteile entspricht, und gelangt so zu folgender Formel:

1) Die Beispiele finden sich bei:

EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 25

$$(4) \quad \bar{S}_i - \bar{S}_j = z_{ij} \sqrt{\sigma_i^2 + \sigma_j^2 - 2r_{ij} \sigma_i \sigma_j}$$

Aufgrund der bekannten p_{ij} -Werte lassen sich die z_{ij} -Werte ermitteln. Die Formel (4) kann dementsprechend nach z_{ij} aufgelöst werden:

$$(5) \quad z_{ij} = (\bar{S}_i - \bar{S}_j) / \sqrt{\sigma_i^2 + \sigma_j^2 - 2r_{ij} \sigma_i \sigma_j}$$

Bei der Formel (5) handelt es sich um den vollständigen Ausdruck des Gesetzes der vergleichenden Urteile.¹⁾

Wenn bei der Technik des Paarvergleichs den Versuchspersonen n Items vorgelegt werden und jedes Item mit jedem anderen Item verglichen wird, so ergibt das - wie bereits erwähnt - eine Gesamtzahl von $n(n-1)/2$ Paaren. Nach dem Beispiel von EDWARDS²⁾ lassen sich auf diese Weise bei der Vorlage von vier Items sechs Gleichungen in der Art der Formel (5) bilden. Die Formel zeigt, daß der Wert der vergleichenden Urteile p_{ij} bzw. der dazugehörigen z_{ij} -Werte eine Funktion der Abstände zwischen den Skalenwerten, $\bar{S}_i - \bar{S}_j$, der Standardabweichungen, σ_i und σ_j , und des Korrelationskoeffizienten r_{ij} ist. Eine Lösung des Systems der sechs Gleichungen des Beispiels ist nicht möglich, weil den sechs bekannten z_{ij} -Werten 14 Unbekannte in Form der vier Skalenwerte, die vier Standardabweichungen und der sechs Interkorrelationen gegenüberstehen.

Um das Gleichungssystem einer Lösung zuzuführen, wird zunächst angenommen, daß alle Standardabweichungen gleich sind: $\sigma_i = \sigma_j = \sigma$. Formel (5) läßt sich dann folgendermaßen vereinfachen:

1) Vgl. THURSTONE, L. L.: A Law of Comparative Judgment, a.a.O., abgedruckt in: L. L. THURSTONE: The Measurement of Values, a.a.O., S. 41

2) Vgl. EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 27 f.

$$(6) \quad z_{ij} = (\bar{S}_i - \bar{S}_j) / \sqrt{2\sigma^2 (1-r_{ij})}$$

Auch in diesem Falle ist eine Lösung nicht möglich, da immer noch mehr Unbekannte als Gleichungen vorhanden sind. Es wird deswegen angenommen, daß alle Interkorrelationen gleich sind; somit kann der Ausdruck umformuliert werden:

$$(7) \quad z_{ij} = (\bar{S}_i - \bar{S}_j) / \sqrt{2\sigma^2 (1-r)}$$

Unter den angenommenen Voraussetzungen ist der Ausdruck unter der Wurzel eine Konstante und kann gleich eins gesetzt werden; daraus folgt:

$$(8) \quad z_{ij} = \bar{S}_i - \bar{S}_j$$

Dieser Ausdruck stellt die einfachste Formulierung des Gesetzes der vergleichenden Urteile dar und wird gewöhnlich als *F a l l V* bezeichnet.¹⁾ Ob die jeweiligen Annahmen, die zur Vereinfachung der ursprünglichen Formel (5) führen, auch tatsächlich berechtigt sind, ist eine Frage empirischer Prüfungen: "The data may or may not be consistent with the model."²⁾ Dieses Problem wird weiter unten noch behandelt; zunächst soll jedoch aufgezeigt werden, wie aufgrund der Ausgangsdaten des Paarvergleichs die Skalenwerte für die einzelnen Items unter Verwendung der Formel (8) errechnet werden können.

1) Vgl.: THURSTONE, L. L.: A Law of Comparative Judgment, a.a.O., abgedruckt in L. L. THURSTONE: The Measurement of Values, a.a.O., S. 41 - 49

In diesem Aufsatz werden auch die weiteren mit eins bis vier bezeichneten Fälle, die jeweils Vereinfachungen des ursprünglichen Modells darstellen, ausführlicher beschrieben.

2) GREEN, B. F.: Attitude Measurement, a.a.O., S. 345

Die Ausgangsdaten bilden eine Häufigkeitsmatrix (F-Matrix): Diese Matrix enthält die Häufigkeiten, mit denen ein Item von den Versuchspersonen in bezug auf die zu messende Einstellung positiver beurteilt worden ist als jedes andere Item. Die Matrix zeigt somit, mit welcher Häufigkeit ein in den Spalten eingetragenes Item positiver als das jeweils in den Zeilen eingetragene Item beurteilt wurde: Die Eintragung f_{ij} ist die Häufigkeit, mit der Item i (es bildet eine Spalte der Matrix) positiver als Item j (es bildet eine Zeile der Matrix) beurteilt wurde. Eine solche F-Matrix hat damit die Form der Tabelle 1.¹⁾

Item	1	2	.	i	.	n
1	f_{11}	f_{21}	.	f_{i1}	.	f_{n1}
2	f_{12}	f_{22}	.	f_{i2}	.	f_{n2}
.	.	.	.	.	.	.
j	f_{1j}	f_{2j}	.	f_{ij}	.	f_{nj}
.	.	.	.	.	.	.
n	f_{1n}	f_{2n}	.	f_{in}	.	f_{nn}

Tabelle 1: Matrix der f_{ij} -Werte (F-Matrix)²⁾

Es wird angenommen, daß ein Vergleich eines Items mit sich selbst $N/2$ ergibt, wenn N die Zahl der Versuchspersonen ist ($f_{ii} = N/2$). Ferner gilt $f_{ji} = N - f_{ij}$.

Aus der F-Matrix wird nun auf dem Wege zur Errechnung der Skalenwerte der Items eine P-Matrix (Tabelle 2) konstruiert,

1) Die Tabellen 1 bis 3 in diesem Abschnitt sind verkürzte und leicht veränderte Wiedergaben der Tabellen 2.1, 2.2 und 2.3 bei:
EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 29, 31 und 32

2) F steht für das englische Wort "Frequency".

wobei lediglich jede Eintragung in der F-Matrix durch N, die Zahl der Versuchspersonen, dividiert wird. Die Eintragungen in der P-Matrix bestehen dann aus den Anteilen an der Gesamtzahl der Urteile, mit denen ein in einer Spalte eingetragenes Item positiver als ein in einer Zeile eingetragenes Item beurteilt wird. Dabei gilt entsprechend der F-Matrix: $p_{ji} = 1 - p_{ij}$. Die Anordnung der Items ist zwar grundsätzlich willkürlich. Meist werden die Items jedoch nach den Spaltensummen der P-Matrix mit dem Item, das die niedrigste Spaltensumme hat, ganz links geordnet. Das erweist sich bei den folgenden Berechnungen der Skalenwerte als vorteilhaft.

Item	1	2	.	i	.	n
1	p_{11}	p_{21}	.	p_{i1}	.	p_{n1}
2	p_{12}	p_{22}	.	p_{i2}	.	p_{n2}
.	.	.	.	.	.	.
j	p_{1j}	p_{2j}	.	p_{ij}	.	p_{nj}
.	.	.	.	.	.	.
n	p_{1n}	p_{2n}	.	p_{in}	.	p_{nn}
Summen	$\sum p_{1.}$	$\sum p_{2.}$	.	$\sum p_{i.}$	.	$\sum p_{n.}$

Tabelle 2: Matrix der p_{ij} -Werte (P-Matrix)¹⁾

Ausgehend von den Eintragungen der P-Matrix kann nun eine Z-Matrix (Tabelle 3) gebildet werden, indem man mit Hilfe einer Normalverteilungstabelle für die p_{ij} -Werte der P-Matrix die entsprechenden z_{ij} -Werte ermittelt und in eine neue Tabelle einträgt. Dabei ist zu beachten, daß

1) P steht für das englische Wort "Proportion".

die Eintragungen oberhalb und unterhalb der Diagonale dieser Matrix, ebenso wie bei der F- und der P-Matrix, nicht unabhängig voneinander sind: Denn wie $f_{ji} = R - f_{ij}$ und $p_{ji} = 1 - p_{ij}$, so gilt $z_{ji} = -z_{ij}$. Wenn also z_{12} die Trennung der Skalenwerte von Item 1 und 2 in der Form $\bar{S}_1 - \bar{S}_2$ wiedergibt, so bedeutet z_{21} dieselbe Trennung in der Form $\bar{S}_2 - \bar{S}_1$, und dieser Wert muß gleich z_{12} sein, jedoch mit umgekehrtem Vorzeichen.¹⁾

Item	1	2	.	i	.	n
1	z_{11}	z_{21}	.	z_{i1}	.	z_{n1}
2	z_{12}	z_{22}	.	z_{i2}	.	z_{n2}
.	.	.	.	.	.	.
j	z_{1j}	z_{2j}	.	z_{ij}	.	z_{nj}
.	.	.	.	.	.	.
n	z_{1n}	z_{2n}	.	z_{in}	.	z_{nn}
Summen	$\sum z_{1.}$	$\sum z_{2.}$	.	$\sum z_{i.}$	.	$\sum z_{n.}$
Mittelwerte	$\bar{z}_{1.}$	$\bar{z}_{2.}$	.	$\bar{z}_{i.}$	.	$\bar{z}_{n.}$
Mittelwerte + Wert mit der größten negativen Abweichung (z.B. $ \bar{z}_{1.} $)	$\bar{z}_{1.} + \bar{z}_{1.} $	$\bar{z}_{2.} + \bar{z}_{1.} $	.	$\bar{z}_{i.} + \bar{z}_{1.} $	.	$\bar{z}_{n.} + \bar{z}_{1.} $

Tabelle 3: Matrix der z_{ij} -Werte (Z-Matrix)

Aus den Werten der so erstellten Z-Matrix können nun unmittelbar die Skalenwerte für die einzelnen Items berechnet werden: Die Eintragungen in den Spalten der Z-Matrix

1) Vgl. EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 31

werden zunächst addiert. Durch Division der so gebildeten Spaltensummen durch die Zahl der Eintragungen in einer Spalte werden anschließend Mittelwerte errechnet. Für den Mittelwert der z-Werte der Spalte i gilt somit:

$$(9) \quad \bar{z}_i = \frac{\sum_{j=1}^n z_{ij}}{n}$$

Der Mittelwert der Z-Werte einer Spalte (z.B. z_{11} bis z_{1n}) ist bei diesem Verfahren der Skalenwert eines Items (z.B. des Items 1), ausgedrückt als Abweichung vom Mittelwert aller Skalenwerte insgesamt. Items, die Skalenwerte mit negativen Vorzeichen haben, bringen somit eine weniger positive Einstellung gegenüber dem betreffenden Einstellungsobjekt zum Ausdruck als der Durchschnitt der Skalenwerte aller Items, und solche Items, deren Skalenwerte ein positives Vorzeichen aufweisen, werden positiver als der Durchschnitt aller Skalenwerte beurteilt.

Weil der Nullpunkt einer so gebildeten Skala als Mittelwert der Skalenwerte aller Items willkürlich gewählt wird, kann er durch Addition einer Konstante zu allen Skalenwerten verschoben werden; denn durch eine solche Transformation wird weder der Abstand zwischen den Skalenwerten noch ihre relative Position auf dem psychologischen Kontinuum verändert.¹⁾ Um alle Skalenwerte durch positive Zahlen auszudrücken, ist es zweckmäßig, den absoluten Skalenwert des Items mit der größten negativen Abweichung allen Skalenwerten hinzuzuaddieren. Dafür ist die letzte Zeile der Tabelle 3 vorgesehen.²⁾

1) Vgl. Abschnitt 32 dieser Arbeit.

2) Eine Arbeitsvereinfachung für die Errechnung der Skalenwerte hat GUILFORD erarbeitet.
Vgl. GUILFORD, J. P.: The Method of Paired Comparisons as a Psychometric Method, Psychol. Rev. 35, 1928, S. 494 - 506
Vgl. auch die ausführliche Darstellung der Technik des Paarvergleichs bei GUILFORD, J. P.: Psychometric Methods, 2. Aufl., New York, Toronto und London 1954, S. 154 - 177; zur Erörterung der Frage der Arbeitsvereinfachung besonders S. 169 ff.

Zur Klärung der Frage, wie weit die empirisch gefundenen Daten mit dem zugrunde liegenden Modell übereinstimmen, kann eine Prüfung auf interne Konsistenz ("internal consistency check") durchgeführt werden.¹⁾ Dabei werden die empirisch festgestellten p_{ij} -Werte mit der Verteilung der Urteile verglichen, wie sie aufgrund der errechneten Skalenwerte erwartet werden müßte. Diese Vorgehensweise stellt gewissermaßen eine Umkehrung des Verfahrens zur Errechnung der Skalenwerte dar. Dazu ist zunächst eine Z'-Matrix aus der Verteilung der für die Items errechneten Skalenwerte abzuleiten.

Items	1	2	3	.	i	.	n
Skalenwerte	0,000	0,602	0,839	.	.	.	.
1	0,000						
2	0,602	-0,602					
3	0,839	-0,839	-0,237				
.	.	.	.				
i	.	.	.	.			
.	.	.	.	.	.		
n	.	.	.	.	.	.	.

Tabelle 4: Matrix der z_{ij} '-Werte (Z'-Matrix)²⁾

Um eine aus den in Tabelle 3 errechneten Skalenwerten abgeleitete Verteilung³⁾ von z_{ij} '-Werten zu erhalten, werden

- 1) Vgl. EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 37 ff.
- 2) Die Tabelle 4 entspricht in verkürzter und leicht geänderter Form der Tabelle 2.7 bei EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 37
- 3) EDWARDS spricht jeweils von "theoretical distribution", "theoretical proportion" usw.
EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 38 ff.
Die Bezeichnung "theoretical" erscheint jedoch in diesem Zusammenhang mißverständlich; in der vorliegenden Arbeit werden deswegen die Ausdrücke "abgeleitete" Verteilung usw. verwendet.

zunächst die Eintragungen auf der linken Seite der Tabelle 4¹⁾ der Reihe nach vom Skalenwert des ersten Items abgezogen und in die erste Spalte der Matrix eingetragen. Entsprechend wird bei den anderen Spalten der Matrix vorgegangen. Auf diese Weise werden die $n(n-1)/2$ Eintragungen unterhalb der Diagonale der Matrix gewonnen. Diese Eintragungen müssen, wenn das Modell den empirischen Daten adäquat ist, als abgeleitete z_{ij} '-Werte den $n(n-1)/2$ unabhängigen z_{ij} -Werten der Tabelle 3 entsprechen. Mit Hilfe einer Normalverteilungstabelle werden die so ermittelten z_{ij} '-Werte in p_{ij} '-Werte transformiert. Diese Werte, die dann den p_{ij} -Werten der Tabelle 2 entsprechen müßten, werden in eine P-Matrix eingetragen. Um einen Vergleich zwischen den Werten der P-Matrix und denen der P'-Matrix zu ermöglichen, werden nun die Eintragungen in der P'-Matrix von den entsprechenden Eintragungen in der P-Matrix abgezogen. Damit werden die Unterschiede zwischen den empirisch vorgefundenen p_{ij} -Werten und den aus den Skalenwerten abgeleiteten p_{ij} '-Werten errechnet; auch diese Werte sind in eine neue Matrix einzutragen. Als Maß für die Unterschiede kann die absolute durchschnittliche Differenz der Werte der beiden Matrices verwendet werden, die EDWARDS nach der Formel

$$(10) \quad AD = \frac{\sum |p_{ij} - p_{ij}'|}{\frac{n(n-1)}{2}}$$

errechnet. Ein Kriterium für die Entscheidung, von welchem Wert an man die absolute durchschnittliche Differenz als zu groß ansieht, als daß die Daten noch mit dem Modell übereinstimmen könnten, gibt es nicht. HEVNER z.B. errechnete einen Wert von 0,024 bei der Skalierung von

1) Es handelt sich dabei um die nach Tabelle 3 errechneten Skalenwerte der Items 1 bis n.

20 Items,¹⁾ SAFFIR einen Wert von 0,031 bei der Skalierung von 25 Items.²⁾ Wie oft in solchen oder ähnlichen Fällen, wird man sich damit aus der Affäre ziehen müssen, daß man die Schaffung von Konventionen zur Entscheidung dieser Frage verlangt.

Eine Komplikation für die Errechnung von Skalenwerten entsteht bei der Technik des Paarvergleichs ebenso wie bei der noch zu erörternden Technik der nachträglich bestimmten Abstände³⁾ dann, wenn die Daten unvollständig sind. Die der Berechnung der Skalenwerte zugrunde gelegten Daten sind dann unvollständig, wenn es nicht zu einer Überlappung der Diskriminationsprozesse kommt, wie sie etwa in der Figur 3 dargestellt wird.⁴⁾ Fehlt eine solche Überlappung - was darin zum Ausdruck kommt, daß alle Versuchspersonen Item i positiver bzw. negativer als Item j beurteilen und die p_{ij} -Werte damit bei 1,00 oder bei 0,00 liegen - so kann man zwar mit großer Sicherheit annehmen, daß Item i eine positivere bzw. negativere Einstellung gegenüber einem Einstellungsobjekt ausdrückt als Item j , eine Grundlage für die Berechnung des Abstandes der Skalenwerte von $\bar{S}_i$ und $\bar{S}_j$ ist jedoch in einem solchen Falle nicht vorhanden, da die z_{ij} -Werte für diese p_{ij} -Werte unbestimmt sind. Eine Überlappung der Diskriminationsprozesse wird vor allem dann nicht vorliegen, wenn zwei Items miteinander verglichen werden, die an entgegengesetzten Enden des psychologischen Kontinuums liegen.⁵⁾

1) Vgl. HEVNER, K.: An Empirical Study of Three Psychophysical Methods, J. Gen. Psychol. 4, 1930, S. 191 - 212.

2) Vgl. SAFFIR, M. A.: A Comparative Study of Scales Constructed by Three Psychophysical Methods, Psychometrika 2, 1937, S. 179 - 193.

3) Vgl. Abschnitt 4313 dieser Arbeit. Vgl. zum folgenden wiederum EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 40 ff.

4) Vgl. S. 98 dieser Arbeit.

5) Vgl. die Skalenwerte $\bar{S}_1$ und $\bar{S}_4$ der Items 1 und 4 in der Figur 3 auf S. 98 dieser Arbeit.

Im allgemeinen werden nicht nur p_{ij} -Werte von 1,00 oder 0,00, sondern auch die in der Nähe liegenden Werte von über 0,98 oder unter 0,02 bei der Berechnung der Skalenwerte außer acht gelassen. Der Grund dafür besteht darin, daß - wie sich anhand einer Normalverteilungstabelle feststellen läßt - die der Differenz zwischen zwei p_{ij} -Werten an den Enden einer Normalverteilung entsprechende Differenz zwischen zwei z_{ij} -Werten weit größer ist als die der gleichen Differenz zwischen zwei p_{ij} -Werten entsprechende Differenz zwischen zwei z_{ij} -Werten im mittleren Bereich einer Normalverteilung. Andere Autoren empfehlen, z_{ij} -Werte von mehr als + 2,0 und weniger als -2,0 nicht zu benutzen; das entspricht p_{ij} -Werten von 0,977 bzw. 0,023.¹⁾

Wenn man also extreme p_{ij} - bzw. z_{ij} -Werte zur Berechnung der Skalenwerte für die Items nicht heranziehen will, so entsteht eine unvollständige Z-Matrix. Damit wird eine etwas andere Berechnungsweise als die bisher beschriebene nötig. Ohne auf die mathematische Grundlage dieser Berechnung im einzelnen einzugehen, lassen sich die Skalenwerte der einzelnen Items aufgrund der Eintragungen in der Matrix der Tabelle 5 ermitteln. Diese Matrix ist eine Umformung einer zunächst aufzustellenden Z-Matrix, die aus den genannten Gründen unvollständig ist. In die Matrix der Tabelle 5 werden die Differenzen zwischen den Werten der angrenzenden Spalten aus der ursprünglichen Z-Matrix eingetragen; diese Differenzen werden addiert (erste Zeile unter der Matrix) und die so gebildeten Summen durch die Zahl der Eintragungen pro Spalte (zweite Zeile unter der Matrix) dividiert. Die auf diese Weise gebildeten Mittelwerte (dritte

1) Vgl. GUILFORD, J. P.: Psychometric Methods, a.a.O., S. 163

Vgl. GREEN, B. F.: Attitude Measurement, a.a.O., S. 345

Zeile unter der Matrix) geben die mittleren Differenzen zwischen den Skalenwerten der Items in den angrenzenden Spalten wieder.

Item	Spaltendifferenzen				
	2 - 1	3 - 2	4 - 3	5 - 4	.
1	1,426	0,000	0,209	-	-
2	1,426	0,065	0,561	0,520	-
3	1,361	0,055	0,292	1,049	.
4	1,019	0,324	0,292	1,076	.
5	-	-0,205	0,205	1,076	.
.	-	-	.	.	.
(1) Summen der Spaltendifferenzen	5,348	0,245	2,373	5,531	.
(2) Zahl der Eintragungen pro Spalte	5	8	9	5	.
(3) Mittelwert	1,170	0,031	0,264	0,691	.
(4) Skalenwerte	$\bar{S}_1=0,000$	$\bar{S}_2=1,170$	$\bar{S}_3=1,201$	$\bar{S}_4=1,465$	$\bar{S}_5=2,156$

Tabelle 5: Matrix der Spaltendifferenzen¹⁾

1) Diese Matrix ist eine verkürzte und leicht veränderte Wiedergabe der Tabelle 212 bei EDWARDS, A. L.: Techniques of Attitude Measurement, a.a.O., S. 48; sie beruht auf den Originaldaten von HILL, R. J.: A Note on Inconsistency in Paired Comparison Judgments, Amer. Soc. Rev. 19, 1953, S. 564 - 566. (Striche (-) bedeuten fehlende Werte, Punkte (.) nicht wiedergegebene Eintragungen aus der ursprünglichen Matrix bei EDWARDS.)

Allgemein gilt demnach:

$$(11) \quad \bar{S}_i - \bar{S}_{(i-1)} = D_{i(i-1)}$$

Dabei ist $D_{i(i-1)}$ die Summe der Differenzen zwischen den angrenzenden Spalten.

Weil die Festsetzung des Nullpunktes der Skala bei dem hier beschriebenen Verfahren willkürlich ist, kann $\bar{S}_1$ als Skalenwert für das erste Item gleich null gesetzt werden und als Skalenwert für das zweite Item gilt dann:

$$(12) \quad \bar{S}_2 = J_{21}$$

Der Skalenwert des dritten Items läßt sich nach der Formel:

$$(13) \quad \bar{S}_3 - \bar{S}_2 + \bar{S}_2 = D_{32} + D_{21} = \bar{S}_3$$

berechnen. Entsprechend wird bei der Berechnung der Skalenwerte für die restlichen Items vorgegangen. (Diese Skalenwerte sind in der vierten Zeile unter der Matrix eingetragen. Für das hier wiedergegebene Beispiel ergibt sich demnach: $\bar{S}_1 = 0,000$; $\bar{S}_2 = 0,000 + 1,170 = 1,170$; $\bar{S}_3 = 1,170 + 0,031 = 1,201$; $\bar{S}_4 = 1,201 + 0,264 = 1,465$ etc.)

Für den Fall einer vollständigen Datenmatrix würde die hier dargestellte Technik zur Errechnung der Skalenwerte zu genau den gleichen relativen Skalenpositionen der Items führen wie die zuerst beschriebene Vorgehensweise. Der Unterschied besteht jedoch darin, daß beim zweiten Verfahren die Trennung zwischen den Skalenwerten benachbarter Items errechnet wird, während beim ersten Verfahren die Skalen-

werte der einzelnen Items als Abweichungen vom Mittelwert aller Skalenwerte gebildet werden.¹⁾

Die Ausführungen dieses Abschnitts machen deutlich, daß die Technik des Paarvergleichs - das gilt auch für die noch zu diskutierende Technik der nachträglich bestimmten Abstände - nicht verwendbar ist, wenn alle Versuchspersonen in ihrem Urteil völlig übereinstimmen. Die Bemerkungen von GREEN, der in diesem Zusammenhang noch einmal die Grundlagen der beiden Verfahren aufzeigt, zeigen das sehr klar; sie seien deshalb als Zusammenfassung zu diesem Punkt etwas ausführlicher zitiert: "The methods of paired comparisons and successive intervals yield interval scales in which the metric is obtained from the variability of judgments about the items. If items were always judged consistently in paired comparisons, so that the proportions were all either zero or unity, the mathematical model of paired comparisons could not be used. ... Happily, judges seldom agree. There is usually sufficient 'error' variance to allow a scale to be constructed. The metric in these scales is essentially a function of the degree of confusion, or overlap, between items. The difference of the scale values of two items may be used to make a precise probability statement about the degree of overlap. The scales are based on variability, but it is a remarkably predictable variability."²⁾

Soviel zur Technik des Paarvergleichs bei der Konstruktion einer Skala. Die Bestimmung von Validität und Reliabilität der auf diese Weise konstruierten Skalen kann mit Hilfe der bereits dargestellten Verfahren erfolgen;³⁾

1) Vgl. EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 46

2) GREEN, J. F.: Attitude Measurement, a.a.O., S. 350

3) Vgl. die Abschnitte 3321 und 3322 dieser Arbeit.

sie sind von der Konstruktion der Skalen unabhängig. Abschließend seien noch einige kurze Bemerkungen zur Frage der Zahl der gemessenen Dimensionen als einem Spezialproblem der Validität sowie dem zu erreichenden Meßniveau angefügt.

Was die Feststellung der Eindimensionalität einer Skala betrifft, so begnügt man sich in der Literatur - sofern man sich überhaupt mit diesem Problem beschäftigt - oftmals mit sehr allgemeinen Feststellungen, wie z.B. den folgenden: "If the data fit the model, the scale is unidimensional by definition. If the data do not fit the model, then either a scale does not exist, or a multidimensional scale is necessary. Unidimensionality has no meaning apart from the context of a particular scaling model."¹⁾ Da aber die Daten in der Forschungspraxis niemals völlig zu dem Modell passen, kann die Frage der Eindimensionalität nicht endgültig entschieden werden; wahrscheinlich muß dementsprechend auch hier die Regelung durch Schaffung einer Konvention erfolgen.

Die Frage nach dem Niveau der Messung wird ebenfalls vielfach zu summarisch behandelt; dafür kann das folgende Zitat von GREEN charakteristisch sein: "The methods of paired comparisons and successive intervals appear to yield interval scales, as does the method of equal-appearing intervals if we make the assumption that equal-appearing intervals are equal."²⁾ Trotz seiner - was das Meßniveau betrifft - im allgemeinen optimistischen Meinung äußert GREEN dann: "An interval scale should have the property that the score distribution is invariant except for a linear transformation, with respect to the particular

1) GREEN, E. F.: Attitude Measurement, a.a.O., S. 363

2) ders., a.a.O., S. 364

Vgl. hierzu auch die Abschnitte 4312 und 4313 dieser Arbeit.

set of items used. It should not be possible to manipulate the score distribution by selecting items with special characteristics. According to this criterion probably none of the scales has equal intervals."¹⁾

Zusammenfassend ist zu sagen: Der Fall V des THURSTONEschen Modells setzt unter anderem Normalverteilung der Diskriminationsprozesse, Gleichheit der Standardabweichungen und Eindimensionalität des psychologischen Kontinuums voraus. Ganz gleich, welchen Test man anwendet, um die Übereinstimmung der empirischen Daten mit den Annahmen des Modells zu prüfen: Immer werden sich mehr oder weniger große Abweichungen ergeben. Es gibt keine Entscheidungskriterien dafür, wann die empirischen Daten die getroffenen Annahmen ausreichend erfüllen; aufgrund der Tests läßt sich lediglich sagen, daß die Annahmen in einem Fall besser als in einem anderen erfüllt werden. Daß die Annahmen in einem konkreten Fall vollständig zutreffen, ist höchst unwahrscheinlich.

-
- 1) GREEN, B. F.: Attitude Measurement, a.a.O., S. 364.
Die Versuche, unter Zugrundelegung des Gesetzes der vergleichenden Urteile sogar Messungen auf Ratioskalenniveau durchzuführen, können an dieser Stelle nicht erläutert werden.
Vgl. hierzu HORST, P.: A Method for Determining the Absolute Affective Value of a Series of Stimulus Situations, J. Educ. Psychol. 23, 1932, S. 411 - 440; - THURSTONE, L. L. und L. V. JONES: The Rational Origin for Measuring Subjective Values, in: L. L. THURSTONE: The Measurement of Values, a.a.O., S. 195 - 210.
Auf dem Verfahren von THURSTONE und JONES bauen auf: HICKS, J. M. und D. T. CAMPBELL: Zero-point Scaling as Affected by Social Object, Scaling Method, and Context, J. Pers. and Soc. Psychol., Bd. 2, Nr. 6, 1955, S. 793 - 800; - SHUFORD, E. H., L. V. JONES und R. D. BOCK: A Rational Origin Obtained by the Method of Contingent Paired Comparisons, Psychometrika 25, 1960, S. 343 - 356.
Vgl. auch die Diskussion bei TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., den Abschnitt: Determination of a Rational Zero Point, S. 194 ff.

Wenn die Skalenwerte für die zu verwendenden Items aufgrund der Expertenurteile errechnet sind, ist die Skalierungsaufgabe abgeschlossen. In der Anwendung der Skala werden die Items den Versuchspersonen, deren Einstellung zu einem bestimmten Einstellungsobjekt gemessen werden soll,¹⁾ mit der Aufforderung vorgelegt, zu jedem Item Ablehnung oder Zustimmung zu äußern. Die Reihenfolge der Items wird zufällig gewählt. Dabei wird von der Annahme ausgegangen,²⁾ daß die so hervorgerufenen Reaktionen der Zustimmung bzw. Ablehnung eine Funktion der Einstellung der Versuchspersonen gegenüber dem betreffenden Einstellungsobjekt sind. Das bedeutet, "that the manner in which an individual responds to these statements, that is, the particular statements that he accepts or rejects would, in turn, enable us to infer something about his location on the same psychological continuum as the one on which the statements have been scaled."³⁾ Zur Bestimmung der Position einer Versuchsperson auf dem psychologischen Kontinuum wird in den meisten Fällen der Median der Skalenwerte der Items verwendet, denen die Versuchsperson zustimmt. Ebenso gut könnte man natürlich - da die vorausgegangenen Operationen ohnehin Intervallskalenniveau voraussetzen - auch das arithmetische Mittel dieser Skalenwerte benutzen. Zudem hat LORGE nur sehr geringe Unterschiede zwischen Positionen, die als Median und solchen, die als arithmetisches Mittel berechnet wurden, festgestellt.⁴⁾

1) In vorliegendem Fall handelt es sich dabei um die am Massenkommunikationsprozeß beteiligten Kommunikatoren und Rezipienten.

2) Dasselbe gilt sinngemäß für alle anderen in dieser Arbeit diskutierten Skalierungsverfahren.

3) EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 47

4) Vgl. LORGE, I.: The Thurstone Attitude Scales: I. Reliability and Consistency of Rejection and Acceptance, J. Soc. Psychol. 10, 1939, S. 187 - 198

Auch andere Verfahren der Anwendung sind erprobt worden: So kann die Versuchsperson z.B. aufgefordert werden, nur das Item anzugeben, das seine Einstellung am besten wiedergibt. Ein Nachteil dieser Vorgehensweise ist jedoch, daß die Versuchsperson dann nur aufgrund des Skalenwertes eines einzigen Items auf dem psychologischen Kontinuum plazierte wird; das kann im Vergleich zu dem bereits beschriebenen Verfahren zu einer Verminderung der Realibilität der Messung führen. EDWARDS¹⁾ schlägt daher vor, die Versuchspersonen zur Auswahl der drei Items aufzufordern, die ihre Einstellung am besten zum Ausdruck bringen, um dann den Median oder den Skalenwert des mittleren Items zur Einstufung der betreffenden Versuchspersonen auf dem psychologischen Kontinuum zu verwenden. An anderer Stelle versucht EDWARDS²⁾, eine mit Hilfe der Technik des Paarvergleichs konstruierte Skala in ihrer Anwendung wie eine kumulative Skala zu behandeln. Zwischen den beiden Vorgehensweisen besteht ein gewisser Unterschied: Bei der Verwendung des Medians oder des arithmetischen Mittels wird die Position einer Versuchsperson auf dem psychologischen Kontinuum festgelegt, auf dem die Items ursprünglich skaliert wurden. Wenn dagegen das zuletzt genannte Verfahren benutzt wird, so werden die Positionen der Versuchspersonen in Form einer linearen Transformation zu dem ursprünglichen psychologischen Kontinuum wiedergegeben.³⁾

1) Vgl. EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 45

2) Vgl. ders.: A Technique for Increasing the Reproducibility of Cumulative Attitude Scales, J. Appl. Psychol. 40, 1955, S. 263 - 265

3) Vgl. ders.: Techniques for Attitude Scale Construction, a.a.O., S. 50

4312 Die Technik der gleicherscheinenden Abstände

Die Technik des Paarvergleichs anzuwenden, ist nur dann zweckmäßig, wenn relativ wenige Items zu skalieren sind. Wie bereits erwähnt, muß jeder Experte schon zur Skalierung von 20 Items 190 vergleichende Urteile abgeben. Diese Schwierigkeit kann umgangen werden, wenn man zur Skalierung der Items die Techniken der gleicherscheinenden Abstände oder der nachträglich bestimmten Abstände benutzt.¹⁾

Bei der THURSTONE und CHAVE²⁾ zugeschriebenen Technik gleicherscheinender Abstände wird zunächst wie beim Paarvergleich eine Anzahl nichtmonotoner Items formuliert. Die Items, die sich auf das Objekt der zu messenden Einstellung beziehen, werden auf einzelne Karten geschrieben.³⁾ In der ursprünglichen Anordnung von THURSTONE und CHAVE werden den Versuchspersonen dann elf Karten vorgelegt, die von links nach rechts geordnet mit den Buchstaben A bis K versehen sind. Die Aufgabe der Versuchspersonen bei der Skalenkonstruktion ist es, die einzelnen Karten mit den Items den mit den Buchstaben versehenen Karten in folgender Weise zuzuordnen: Der A-Karte sind die Items, die die negativste Einstellung gegenüber dem Einstellungsobjekt ausdrücken, zuzuteilen. Dementsprechend entfallen auf die K-Karte die Items, in denen die positivste Einstellung

1) So empfiehlt GREEN diese Techniken von 20 oder mehr Items an.

Vgl. GREEN, B. F.: Attitude Measurement, a.a.O., S. 347

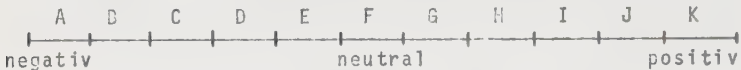
2) Vgl. THURSTONE, L. L. und E. J. CHAVE: The Measurement of Attitude, a.a.O.

SCHEUCH bemerkt, daß das Verfahren wahrscheinlich zuerst von THORNDIKE 1900 entwickelt wurde. "Dennoch gilt L. L. THURSTONE als Entdecker des Verfahrens für die Sozialforschung."

SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, a.a.O., S. 356

3) Vgl. hierzu die Darstellung bei EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 83ff.

zum Ausdruck kommt. Die in der Mitte liegende F-Karte wird als neutrale Karte gekennzeichnet; ihr sollen die Items zugeteilt werden, die nach Meinung der Versuchspersonen weder eine positive noch eine negative Einstellung wiedergeben. Die Karten E bis A repräsentieren bei dieser Einteilung eine zunehmend negativere Einstellung, die Karten G bis K eine zunehmend positivere Einstellung.



Das Kontinuum gleicherscheinender Intervalle
von THURSTONE

Jede Versuchsperson ordnet nun jedes Item einer der elf vorgegebenen Kategorien zu, die in der beschriebenen Weise ein Kontinuum von positiv (A) über neutral (F) bis negativ (K) bilden. Über die Zahl der benötigten Versuchspersonen zur Einstufung der Items bestand zunächst keine Einigkeit. THURSTONE und CHAVE setzten 300 Versuchspersonen zur Einordnung von 130 Items bei der Konstruktion ihrer Skala zur Messung der Einstellung gegenüber der Kirche ein. In späteren Untersuchungen¹⁾ wurde die Zahl der Versuchspersonen drastisch herabgesetzt, z.B. bis auf zwei Gruppen von je 15 Versuchspersonen²⁾. Dabei wurden fast vollständige Korrelationen bei der Errechnung der Skalenwerte der Items bei kleinen und bei großen Gruppen von Versuchspersonen festgestellt.

- 1) Vgl. NYSTROM, G. H.: The Measurement of Filipino Attitudes toward America by the Use of the Thurstone Technique, J. Soc. Psychol., 4, 1933, S. 249 - 252
Vgl. UHRBROCK, R. S.: Attitudes of 4430 Employees, J. Soc. Psychol., 5, 1934, S. 365 - 377
Vgl. FERGUSON, L. W.: The Requirements of an Adequate Attitude Scale, Psychol. Bull., 36, 1939, S. 665 - 673
Vgl. EDWARDS, A. L. und K. C. KENNEY: A Comparison of the Thurstone and Likert Techniques of Attitude Scale Construction, J. Appl. Psychol. 30, 1946, S. 72 - 83; abgedruckt in: W. FISHBEIN (Hrsg.): Readings in Attitude Theory and Measurement, New York, London und Sydney 1967, S. 249 - 256
- 2) Vgl. ROSANDER, A. C.: The Spearman-Brown Formula in Attitude Scale Construction, J. Exp. Psychol. 19.

Außer der mittleren und den beiden extremen Positionen erfolgt keine weitere Kennzeichnung der dazwischen liegenden Karten. Damit soll erreicht werden, daß den Versuchspersonen die Abstände zwischen den einzelnen Karten, die als Intervalle auf dem Kontinuum von positiv bis negativ behandelt werden, gleich erscheinen. Das ist die Grundannahme, auf der das gesamte Verfahren aufbaut, aber: "It is not a particularly good assumption".¹⁾ Denn die Technik bietet keine Möglichkeit der Prüfung, ob diese Annahme auch tatsächlich gegeben ist. Eine Entscheidung darüber, wieweit die Abstände zwischen den einzelnen Intervallen A bis K wirklich als gleich groß wahrgenommen werden, muß somit auf bloßen Vermutungen beruhen. Ein Nachweis dafür, daß mit diesem Verfahren eine Intervallskala konstruiert wird, kann - trotz des Namens dieser Technik - nicht erbracht werden. Ergebnisse empirischer Untersuchungen sprechen im Gegenteil dafür, daß z.B. die Versuchspersonen im mittleren Bereich des Kontinuums feinere Unterscheidungen machen als an den beiden Endpunkten. HEVNER hat das in einer Untersuchung nachgewiesen, in der sie eine Skala mit Hilfe der Techniken gleicherscheinender Abstände und des Paarvergleiches konstruierte.²⁾ Die Beziehungen zwischen den Punktwerten der Items der beiden Skalen waren fast linear, außer an den Extremen, wo die Punktwerte der Items, die nach dem Verfahren gleicherscheinender Abstände errechnet waren, näher zusammenrückten.

Die Ergebnisse dieser Untersuchung werden durch folgende Überlegung gestützt: Wenn man für die Zuordnung der Items zu den einzelnen Intervallen eine Normalverteilung annimmt, so ist eine solche Verteilung nur für die mittleren Be-

1936, S. 486 - 495

1) GREEN, F. B.: Attitude Measurement, a.a.O., S. 347

2) Vgl. HEVNER, K.: An Empirical Study of Three Psychophysical Methods, a.a.O.

reiche des Kontinuums denkbar. Da das Kontinuum nach den Extrempunkten A und K "abgeschnitten" wird, ist in diesen Bereichen eine normalverteilte Zuordnung der Items unmöglich. Unabhängig davon, wie beim Verfahren gleicherscheinender Abstände die Punktwerte für die einzelnen Items berechnet werden, muß es somit bei dieser Berechnung zu einer Verschiebung zum Mittelpunkt des Kontinuums kommen.¹⁾ Das bedeutet, daß man für die Hypothesenbildung davon ausgehen muß, daß sich mit Hilfe der Technik gleicherscheinender Abstände lediglich Instrumente konstruieren lassen, die auf dem Niveau von Ordinalskalen messen. Die diesbezügliche Schwäche der Technik wird besonders augenfällig bei der Berechnung der Punktwerte für die einzelnen Items, wie THURSTONE und CHAVE sie durchführen; denn dabei wird die genannte zweifelhafte Annahme gleicher Abstände verwendet.

Wenn die Abstände zwischen den einzelnen Kategorien von den Versuchspersonen gleich wahrgenommen werden, können die Kategorien A bis K mit den Zahlen 1 bis 11 durchnummeriert werden,²⁾ wobei man die Zahlenreihe als Intervallskala interpretiert und die einer solchen Skala zugrunde liegenden Postulate auch für die Kategorien A bis K als gegeben annimmt. Angemessener wäre es jedoch, von vornherein nur die Konstruktion einer Ordinalskala anzustreben und für die Zahlenreihe 1 bis 11 und damit auch die Kategorien A bis K lediglich das Vorliegen der Postulate einer solchen Skala anzunehmen. Das erscheint um so sinnvoller, als bestimmte Formen der nunmehr vorzunehmenden Operationen

1) Vgl. ATTNEAVE, F.: A Method of Graded Dichotomies for the Scaling of Judgments, Psychol. Rev., 56, 1949, S. 334

2) Möglich ist natürlich auch die Benutzung einer anderen Zahlenreihe, z.B. von -5 über 0 als neutralen Punkt bis + 5.

zur Ermittlung der Punktwerte und der geeignetsten Items für eine Skala (S- und Q-Werte) bereits auf Ordinalniveau möglich sind.

Zur Berechnung des Punktwertes der Items werden die Daten, nachdem alle Versuchspersonen jedes Item einer Kategorie zugeordnet haben, in Form der folgenden Tabelle dargestellt:¹⁾

Item	Kategorien						Skalenwert	Q-Wert
	A 1	B 2		J 10	K 11			
f								
1 p								
cp								
f								
2 p								
cp								
.								
.								
.								

Tabelle 6: Daten zur Berechnung der S- und Q-Werte

Die Tabelle enthält für jedes Item drei Zeilen. In der ersten Zeile (f) sind die Häufigkeiten eingetragen, mit denen ein Item der betreffenden Kategorie zugeordnet wurde. In der zweiten Zeile sind diese Häufigkeiten in Anteile an der Zahl der Gesamturteile umgeformt (p).²⁾ Die dritte

1) Vgl. EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., Tabelle 4.1, S. 87

2) Im folgenden wird nur noch von Anteil("proportion") gesprochen.

Zeile schließlich gibt die kumulativen Anteile ($\sum p_u$) wieder.

Die Werte dieser Tabelle dienen als Grundlage für die Erreichung der Skalierungsaufgabe, die darin besteht, eine Reihe von Items zu finden, deren Punktwerte gleichmäßig über das Einstellungskontinuum verteilt sind und bei denen die Streuung in der Verteilung der Urteile der Versuchspersonen möglichst gering ist. Als Punktwert eines Items kann der Median der Verteilung der Urteile für das betreffende Item verwendet werden. Er läßt sich nach der folgenden Formel errechnen:

$$(1) S = g + \left(\frac{0,50 - \sum p_u}{p_i} \right) i$$

Dabei ist:

S = Median als Punktwert eines Items

g = Untergrenze des Intervalls, in dem der Median liegt.¹⁾

$\sum p_u$ = Summe der Anteile unter dem Intervall, in dem der Median liegt

p_i = Anteil innerhalb des Intervalls, in dem der Median liegt

i = Intervallbreite (wird hier mit 1 angesetzt).

1) Bei der Berechnung werden Intervallgrenzen von 0,5 einer Einheit unter bis 0,5 einer Einheit über der dem betreffenden Intervall zugeordneten Zahl zugrunde gelegt. So reichen z.B. die Intervallgrenzen der Kategorie 2 mit der zugehörigen Zahl 2 von 1,5 bis 2,5.

Um eine Berechnung nach dieser Formel anhand der Werte der Tabelle 6 durchführen zu können, muß eine Intervallskala bestehen. Geht man jedoch davon aus, daß lediglich eine Ordinalskala vorliegt, so kann aufgrund der dritten Tabellenzeile mit den kumulativen Anteilen (cp) für jedes Item festgestellt werden, in welche Kategorie der Median der Verteilung der Urteile fällt (z.B. in die mit 8 bezeichnete Kategorie K). Damit wäre eine weniger feine, jedoch dem Ordinalskalenniveau angemessene Berechnungsweise gewählt. Es müßte jeweils durch empirische Prüfung entschieden werden, wieweit sich mit ihrer Hilfe das erste Skalierungsziel erreichen läßt: die Gewinnung von Items, die gleichmäßig über das gesamte Einstellungskontinuum von negativ bis positiv verteilt sind.

Die zweite Aufgabe besteht in der Berechnung der Streuung der Urteile für jedes Item. THURSTONE und CHAVE verwenden als Maß für die Streuung die interquartile Spannweite Q ("interquartile range"), die sich auf die mittleren 50 % der Urteile erstreckt und nach der Formel

$$(2) \quad Q = C_{75} - C_{25}$$

berechnet wird. Dabei ist

C_{75} = das 75. Zentil

C_{25} = das 25. Zentil¹⁾.

Diese Berechnung setzt wiederum das Vorliegen einer Intervallskala voraus. Man kann sie jedoch auch in Anlehnung an die oben vorgeschlagene Berechnungsweise des Medians so durchführen, daß man aufgrund der Werte der Tabelle feststellt, in welche Kategorien C_{75} und C_{25} fallen, z.B. in

1) Die Berechnung der Werte C_{75} und C_{25} erfolgt durch Umformung der Formel (1), die in der dargestellten Art zur Berechnung des 50. Zentils (C_{50} gleich dem Median) dient.

die mit 10 bezeichnete Kategorie J und in die mit 6 bezeichnete Kategorie F. Das wäre eine dem Ordinalskalenniveau angemessene Vorgehensweise; zu prüfen ist in diesem Falle jedoch wiederum, ob diese Berechnungsweise zur Ermittlung der Items mit den geringsten Streuungswerten fein genug ist, denn nur diese werden in die endgültige Skala zur Messung einer Einstellung aufgenommen.¹⁾

Das Ziel der Skalierung ist somit die Bildung einer Skala von 20 bis 22 Items,²⁾ deren Punktwerte gleichmäßig über das psychologische Kontinuum von positiv bis negativ verteilt sind und deren Q-Werte möglichst gering sind. Werden für zwei Items gleiche oder nahe beieinander liegende Punktwerte errechnet, so wird das Item mit dem geringeren Q-Wert für die endgültige Skala ausgewählt. Wenn genügend Items zur Verfügung stehen, ist es leicht, zwei Skalen von der beschriebenen Art zur Messung ein und derselben Einstellung zu bilden, um so Reliabilitätstests mit Hilfe des Verfahrens der äquivalenten Formen durchführen zu können.³⁾

In der Anwendung einer auf diese Weise konstruierten Skala werden die einzelnen Items einer Person zur Annahme oder Ablehnung vorgelegt. Der Punktwert einer Person wird dann als Median oder arithmetisches Mittel der Punktwerte der

1) Es besteht auch die Möglichkeit, die S- und Q-Werte graphisch zu ermitteln. Vgl. dazu EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 90 - 92 und die dort zitierte Literatur.

2) Vgl. ders., a.a.O., S. 92
SCHEUCH spricht von "14 oder 22 Aussagen".
SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, a.a.O., S. 357

3) Nach EDWARDS werden bei dieser Vorgehensweise Reliabilitätskoeffizienten von 0,85 und mehr erzielt. Vgl. EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 94

von ihr akzeptierten Items errechnet. "Hierdurch wird allerdings der Gesamtpunktwert nicht nur - wie beabsichtigt - eine Funktion der Art der akzeptierten Aussagen, sondern gleichfalls der Zahl der Akzeptierungen, wodurch viele positive Reaktionen den Effekt haben, die Wirkung der Annahme einer extremen Aussage im Punktwert zu vermindern, und umgekehrt."¹⁾ GUILFORD schlägt deshalb vor, die Zahl der Antwortmöglichkeiten auf drei oder fünf zu beschränken, um so eine Konzentrierung auf die dem Durchschnitt der betreffenden Person am nächsten kommenden Items und deren Punktwerte zu erreichen.²⁾

Neben der von THURSTONE und CHAVE vorgeschlagenen Berechnungsweise für die S- und Q-Werte der einzelnen Items, die darum problematisch ist, weil sie dem Ordinalskalenniveau nicht angemessen ist, muß noch ein weiterer Mangel der Technik gleicherscheiner Abstände erörtert werden: Die Urteile der Versuchspersonen können, wenn sie bei der Konstruktion der Skala als Experten zur Einstufung der Items aufgefordert sind, in einer unerwünschten Weise beeinflusst werden. Vor allem zwei Faktoren können eine solche Beeinflussung verursachen: a) die Art, in der die Versuchspersonen veranlaßt werden, ihre Urteile abzugeben, und b) die Einstellungen, die die Versuchspersonen selbst zu dem Einstellungsobjekt haben, auf das sich die von ihnen zu beurteilenden Items beziehen.

Abweichend von der ursprünglichen Vorgehensweise THURSTONES und CHAVES wurden verschiedene weitere Verfahren für die Abgabe der Urteile von Versuchspersonen entwickelt. Weil bei den einzelnen hierzu durchgeführten Untersuchungen dieselben

1) SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, a.a.O., S. 357

2) Vgl. GUILFORD, J. P.: Psychometric Methods, New York, Toronto und London 1954, S. 459
Vgl. hierzu auch die entsprechenden Ausführungen im vorausgehenden Abschnitt dieser Arbeit.

Items zum Teil zweimal auf unterschiedliche Weise beurteilt wurden, lassen die Ergebnisse dieser Studien u.a. eine Einschätzung des Einflusses verschiedener Urteilsweisen zu.¹⁾ SEASHORE und HEVNER²⁾ schrieben neben die einzelnen zu beurteilenden Items die Zahlenreihe von 1 bis 11 und veranlaßten die Versuchspersonen, die Zahl anzukreuzen, die nach ihrer Ansicht die Position des betreffenden Items auf dem Einstellungskontinuum wiedergibt. DALLIN und FARNSWORTH³⁾ benutzten ein graphisches Verfahren, bei dem die Versuchspersonen die Position eines Items auf dem Einstellungskontinuum durch Ankreuzen der entsprechenden Stelle auf einer Geraden bestimmter Länge angaben. Erst anschließend wurde diese Gerade in gleiche Einheiten eingeteilt, um so die zu einem Item gehörige Kategorie bestimmen und die S- und Q-Werte berechnen zu können. Bei EDWARDS und KILPATRICK⁴⁾ wurden die Versuchspersonen zunächst veranlaßt, die zu beurteilenden Items drei Kategorien zuzuordnen: einer positi-

1) Hauptziel dieser Untersuchungen ist es meist, ein Verfahren zu erproben, das gegenüber der ursprünglichen Technik zu Zeit- und Kostenersparnissen führt.

2) Vgl. SEASHORE, R. W. und K. HEVNER: A Time-saving Device for the Construction of Attitude Scales, J. Soc. Psychol. 4, 1933, S. 366 - 372

Dieselbe Verfahrensweise wurde auch von EDWARDS und KENNEY zur Skalierung von 129 ursprünglich von THURSTONE und CHAVE skalierten Items benutzt. Für die Skalenwerte bei diesen beiden Skalierungsversuchen ergaben sich Korrelationen von 0,95.

Vgl. EDWARDS, A. L. und K. C. KENNEY: A Comparison of the Thurstone and Likert Techniques of Attitude Scale Construction, a.a.O.

3) Vgl. DALLIN, M. und P. R. FARNSWORTH: A Graphic Rating Method for Determining the Scale Values of Statements in Measuring Social Attitudes, J. Soc. Psychol., 13, 1941, S. 323 - 327

4) Vgl. EDWARDS, A. L. und F. P. KILPATRICK: A Technique for the Construction of Attitude Scales, a.a.O.

ven, einer neutralen und einer negativen Kategorie. Jede dieser drei Kategorien wurde anschließend in drei weitere unterteilt; so wurde z.B. die Gruppe der positiven Items in eine Gruppe der am meisten positiven, eine der am wenigsten positiven und eine mittlere Gruppe unterteilt. Auf diese Weise wollen EDWARDS und KILPATRICK den Versuchspersonen den Überblick über alle zu beurteilenden Items, die Art der zu fällenden Urteile und das gesamte psychologische Kontinuum erleichtern.¹⁾ WEBB hat eine Untersuchung durchgeführt, bei der die Zuordnung der Items mit Hilfe von IBM-Maschinen vorgenommen wurde.²⁾ Die Ergebnisse der zitierten Untersuchungen lassen aufgrund der erzielten hohen Reliabilitätskoeffizienten die Vermutung zu, daß die Festlegung der Skalenwerte der Items nicht von den jeweils verwendeten Urteilsverfahren beeinflusst wird.

Andererseits zeigten sich jedoch in der Berechnung der Q-Werte erhebliche Unterschiede zwischen den neueren Untersuchungen von EDWARDS und KENNEY³⁾ und WEBB⁴⁾ und den von THURSTONE und CHAVE durchgeführten Studien. EDWARDS führt dazu aus, daß unklar sei, ob diese Differenzen auf die unterschiedlichen Urteilsverfahren oder auf mögliche Einflüsse der zwischen den Untersuchungen liegenden Zeitspanne zurückzuführen seien.⁵⁾ Jedenfalls zeigen die Er-

1) Vgl. EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 97

2) Vgl. WEBB, S. C.: A Generalized Scale for Measuring Interest in Science Subjects, Educ. Psychol. Measurement 11, 1951, S. 456 - 469

3) Vgl. EDWARDS, A. L. und K. C. KENNEY: A Comparison of the Thurstone and Likert Techniques of Attitude Scale Construction, a.a.O.

4) Vgl. WEBB, S. C.: A Generalized Scale for Measuring Interest in Science Subjects, a.a.O.

5) Vgl. EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 99

gebnisse der Untersuchungen, auf welch unsicherem Boden selbst ein in den Sozialwissenschaften so häufig verwendetes Verfahren wie das der gleicherscheinenden Abstände im Vergleich zu den meisten der in den Naturwissenschaften benutzten Techniken steht.

Das wird noch deutlicher, wenn man den zweiten wichtigen Faktor in die Betrachtung einbezieht, der die Urteile der Versuchspersonen beeinflussen kann. Die Technik gleicherscheinender Abstände beruht - wie gesagt - auf der Annahme, daß die Versuchspersonen, die als Experten eingesetzt werden, die Beurteilung der Items unabhängig von ihrer eigenen Einstellung zu dem in Frage stehenden Einstellungsobjekt vornehmen. Wie problematisch jedoch diese Unterstellung einer sogenannten "objektiven Einstufung"¹⁾ ist, die im Widerspruch zu den Ergebnissen sozialpsychologischer Forschung steht, zeigen die im Anschluß an die Entwicklung des Verfahrens durchgeführten diesbezüglichen Untersuchungen. An dieser Stelle soll ein möglichst geraffter Überblick über die Forschungen zu diesem Problem, die bis in die jüngste Zeit fortgesetzt wurden, gegeben werden. Wenn dieser Überblick trotzdem relativ umfangreich ist, so liegt das an der großen Fülle der Literatur gerade zu diesem Fragenkreis. Denn die Technik der gleicherscheinenden Abstände erfreut sich wegen ihrer vergleichsweise leichten Anwendbarkeit trotz fragwürdiger Voraussetzungen nach wie vor großer Beliebtheit. Zudem ermöglicht eine Übersicht über die bisherigen Forschungsergebnisse einen Vergleich der Technik gleicherscheinender Abstände mit anderen "Judgment"-Techniken.

1) Vgl. SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, a.a.O., S. 357

Zunächst schienen die Untersuchungen die genannte Unterstellung zu rechtfertigen. HINCKLEY¹⁾ setzte zur Beurteilung von Items, die Einstellungen gegenüber Negern erfassen sollten, drei Gruppen von Versuchspersonen ein, von denen er annehmen konnte, daß sie in der Extremität²⁾ ihrer Einstellungen gegenüber Negern variierten: weiße Südstaatler (negative Einstellung), weiße Nordstaatler (positive Einstellung) und Neger. Diese Gruppen hatten 114 Items nach dem beschriebenen Verfahren in elf Kategorien einzustufen. Die Punktwerte für die Items wurden dann für jede Gruppe getrennt berechnet; sie korrelierten bei den beiden Gruppen von Nord- und Südstaatlern und Negern mit 0,93. HINCKLEY schloß daraus, daß die Beurteilung der Items unabhängig von der Einstellung der Versuchspersonen erfolgt war. Dieselben Ergebnisse wurden aufgrund ähnlicher Versuchsanordnungen in verschiedenen anderen Untersuchungen erzielt.³⁾ Jeweils wurden die Punktwerte der Items für einzelne Gruppen mit unterschiedlichen Einstellungen zu einem Einstellungsobjekt getrennt berechnet und anschließend korreliert. Die Korrelationen sind fast vollständig (um 0,98). Es fragt sich allerdings, ob die jeweils gewählten Versuchsanordnungen und Berechnungsweisen eine adäquate

-
- 1) Vgl. HINCKLEY, E. D.: The Influence of Individual Opinion on Construction of an Attitude Scale, J. Soc. Psychol. 3, 1932, S. 283 - 296
 - 2) Der Terminus Extremität wird hier wiederum gleichbedeutend mit Intensität im oben beschriebenen Sinne verwendet.
 - 3) Vgl. BEYLE, H. C.: A Scale for the Measurement of Attitude toward Candidates for Elective Governmental Office, Amer. Polit. Sci. Rev. 26, 1932, S. 527 - 544
Vgl. FERGUSON, L. W.: The Influence of Individual Attitudes on Construction of an Attitude Scale, J. Soc. Psychol. 6, 1935, S. 115 - 117
Vgl. PINTER, R. und G. FORLANO: The Influence of Attitude upon Scaling of Attitude Items, J. Soc. Psychol. 8, 1937, S. 39 - 45
Vgl. EYSENCK, H. J. und S. CROWN: An Experimental Study in Opinion-Attitude Methodology, Int. J. Opin. und Attitude Res. 3, 1949, S. 47 - 86

Prüfung der These THURSTONEs und CHAVEs ermöglicht haben. So stellen EDWARDS und KENNEY fest: "When social approval or disapproval attaches to a favorable or unfavorable attitude toward an issue, different scale values might result from groups with differing attitudes ... The research so far ... also neglects the related problem of ego-involved attitudes and the bearing they might have upon scale values of items."¹⁾

In einer Kritik der Studie von HINCKLEY vermuten HOVLAND und SHERIF²⁾, daß die Anwendung des THURSTONE-Kriteriums zur Ausscheidung nicht sorgfältig arbeitender Versuchspersonen³⁾ dazu geführt hat, daß die Urteile derjenigen Versuchspersonen mit den extremsten Einstellungen gegenüber Negern gar nicht zur Berechnung der Punktwerte der Items herangezogen wurden und auf diese Weise die hohen Korrelationen zwischen den für die verschiedenen Gruppen mit unterschiedlichen Einstellungen gegenüber Negern errechneten Punktwerte zustande kommen konnten. HOVLAND und SHERIF formulierten drei Hypothesen über den Zusammenhang zwischen der Extremität der Einstellung von Versuchspersonen und der von ihnen vorzunehmenden Einstufung von Items nach der Technik der gleicherscheinenden Abstände:⁴⁾

-
- 1) EDWARDS, A. L. und K. C. KENNEY: A Comparison of the Thurstone and Likert Techniques of Attitude Scale Construction, a.a.O., S. 82
 - 2) Vgl. HOVLAND, C. I. und M. SHERIF: Judgmental Phenomena and Scales of Attitude Measurement: Item Displacement in Thurstone Scales. J. Abnorm. Soc. Psychol. 47, 1952, S. 822 - 832, abgedruckt in: M. FISHBEIN (Hrsg.): Readings in Attitude Theory and Measurement, a.a.O., S. 236 - 248
 - 3) THURSTONE und CHAVE eliminierten alle Versuchspersonen, die bei der Konstruktion ihrer Skala zur Messung der Einstellung gegenüber der Kirche 30 oder mehr der insgesamt 130 zu beurteilenden Items in eine der elf vorgegebenen Kategorien einordneten. Das entspricht einem Prozentsatz von mindestens 23 % aller Items in einer Kategorie.
 - 4) Vgl. HOVLAND, C. I. und M. SHERIF: Judgmental Phenomena and Scales of Attitude Measurement: Item Displacement in Thurstone Scales, a.a.O., S. 824

- a) Versuchspersonen mit extrem positiven oder extrem negativen Einstellungen zeigen die Tendenz, ihre Einstufung der Items auf eine geringe Zahl von Kategorien zu konzentrieren.
- b) Versuchspersonen mit extremer Einstellung und starker Ich-Beteiligung ("strong personal involvement") zeigen die Tendenz, sehr feine Unterscheidungen auf ihrer Seite der Skala zu machen ("raised threshold of acceptance") und die Items auf der anderen Seite der Skala pauschal zusammenzufassen ("lowered threshold of rejection").
- c) Eine Verschiebung des Punktwertes der Items wird im "neutralen" Bereich stärker hervortreten als für eindeutig positiv oder negativ formulierte Items.¹⁾

Bei ihrer Untersuchung verwendeten HOVLAND und SHERIF dieselbe Vorgehensweise und Versuchsanordnung wie HINCKLEY und legten ihren Versuchspersonen auch dieselben Items zur Beurteilung vor. Die Ergebnisse bestätigten die genannten Hypothesen und zeigten auch, daß das von THURSTONE und HINCKLEY angewendete Kriterium zur Eliminierung nicht sorgfältig arbeitender Versuchspersonen für die hohen Korrelationen zwischen den Punktwerten der Items, die für verschiedene Gruppen mit unterschiedlichen Einstellungen getrennt errechnet wurden, verantwortlich ist. Werden auch die Urteile der Personen, die mehr als 30 Items in eine Kategorie einordnen, zur Errechnung der Punktwerte der Items herangezogen, so ergeben sich die in den Hypothesen b) und c) behaupteten Verschiebungen.²⁾

1) Wie HOVLAND und SHERIF im einzelnen zur Aufstellung dieser Hypothesen gelangen, soll hier nicht erläutert werden.

2) Vgl. vor allem die graphische Darstellung, die sich bei HOVLAND und SHERIF auf S. 829 befindet.

In einer späteren Untersuchung beobachteten KELLEY, HOVLAND, SCHWARTZ und ABELSON¹⁾ dieselben Itemverschiebungen wie vorher HOVLAND und SHERIF. Wenn dieselben Items nach der Technik des Paarvergleichs skaliert wurden, zeigten sich diese Verschiebungen nicht. Das bedeutet, daß die Itemverschiebungen eine der Technik der gleicherscheinenenden Abstände eigentümliche Erscheinung ist, die nicht beobachtet wird, wenn Items mit Hilfe eines anderen Verfahrens skaliert werden.

Weitere Faktoren, die eine Verschiebung des Punktwertes der Items verursachen können, wurden in anderen Studien untersucht; damit bestehen alternative Erklärungsmöglichkeiten für die Itemverschiebungen. PROTHRO²⁾ gelangt zu Ergebnissen, die von HOVLAND und SHERIF abweichen: Er stellte bei den von ihm eingesetzten Gruppen von Versuchspersonen keine Itemverschiebungen fest und erklärt das damit, daß in der Untersuchung von HOVLAND und SHERIF eine Gruppe der Versuchspersonen, die Neger, die Möglichkeit der persönlichen Identifikation mit dem Einstellungsobjekt bei der Beurteilung von Items hatten, weil sich diese auf ihre eigene soziale Gruppe bezogen. Bei den von PROTHRO gewählten Versuchspersonen bestand keine solche Identifikationsmöglichkeit. Damit liegt die Vermutung nahe, daß die Frage, wieweit sich die Versuchspersonen mit einem Einstellungsobjekt, auf das sich die zu beurteilenden Items beziehen, identifizieren können oder nicht, ein Faktor ist, der die Urteile beeinflussen kann. In einer Untersuchung von GRANNEBERG³⁾ zur Einstellung gegenüber der Religion

1) Vgl. KELLEY, H. H., C. I. HOVLAND, M. SCHWARTZ und R. T. ABELSON: The Influence of Judges' Attitudes in Three Methods of Attitude Scaling, J. Soc. Psychol. 42, 1955, S. 147 - 147 - 158

2) Vgl. PROTHRO, E. T.: The Effect of Strong Negative Attitudes on the Placement of Items in a Thurstone Scale, J. Soc. Psychol. 41, 1955, S. 11 - 18
Vgl. auch ders.: Personal Involvement and Item Displacement on Thurstone Scales, J. Soc. Psychol. 45, 1957, S. 191 - 196

3) Vgl. GRANNEBERG, R. T.: The Influence of Individual Attitude and Attitude-Intelligence upon Scale Values

zeigte sich, daß die Einstellungen der Versuchspersonen gegenüber der Religion und ihr Intelligenzniveau sich gegenseitig beeinflussen und gemeinsam auf die Beurteilung von Item einwirken, die sich auf die Religion beziehen. GRANDEBERG nimmt an, daß die von ihm beobachteten Itemverschiebungen nicht allein auf die unterschiedliche Ich-Beteiligung¹⁾ der Versuchspersonen, sondern auf die Wechselwirkung der beiden Variablen Ich-Beteiligung und Intelligenzniveau zurückzuführen ist. FEHRER²⁾ zeigt in ihrer Untersuchung, daß der Zusammenhang, in dem ein Item zu den anderen Items steht, die skaliert werden sollen, auch ein Faktor ist, der die Berechnung des Punktwertes eines Items beeinflussen kann.

In der Folgezeit wurde die Diskussion der Probleme der Itemverschiebung und der Faktoren, die eine solche Verschiebung bewirken, vor allem im Journal of Abnormal and Social Psychology bzw. im Journal of Personality and Social Psychology fortgesetzt. Die Einzelheiten dieser Diskussion sollen hier nicht weiter verfolgt werden, da die grundsätzlichen theoretischen und praktischen Schwierigkeiten bei der Technik gleicherscheinender Abstände

of Attitude Items, Amer. Psychol. 10, 1955, S. 330 f.

- 1) Der Ausdruck Ich-Beteiligung steht hier für die englischen Termini: "ego involvement", "attitudinal involvement" oder "personal involvement". Zum Problem der Ich-Beteiligung vgl. in diesem Zusammenhang auch: WARD, C. D.: Ego Involvement and the Absolute Judgment of Attitude Statements, J. Pers. Soc. Psychol. 2, 1965, S. 202 - 208; - ders.: Attitude and Involvement in the Absolute Judgment of Attitude Statements, J. Pers. Soc. Psychol. 4, 1966, S. 465 - 476
- 2) Vgl. FEHRER, E.: Shifts in Scale Values of Attitude Statements as a Function of the Composition of the Scale, J. Exp. Psychol. 42, 1955, S. 179 - 186

bereits aus den bisherigen Ausführungen hervorgehen dürfen.¹⁾ Abschließend ist noch zu bemerken, daß die mit Hilfe dieser Technik konstruierten Skalen mit allen oben erläuterten Vorgehensweisen der Validitätsbestimmung geprüft werden können; Reliabilitätsprüfungen sind ebenfalls mit allen drei genannten Verfahren möglich.

-
- 1) In den vergangenen Jahren ist in der genannten Zeitschrift eine Fülle von Aufsätzen zu diesen Fragen erschienen. Es seien lediglich die Artikel von MANIS und UPSHAW angeführt:
Vgl. MANIS, M.: The Interpretation of Opinion Statements as a Function of Recipient Attitude, J. Abn. Soc. Psychol. 60, 1960, S. 340 - 344; - ders.: The Interpretation of Opinion Statements as a Function of Message Ambiguity and Recipient Attitude, J. Abn. Soc. Psychol. 63, 1961, S. 76 - 81; - ders.: The Interpretation of Opinion Statements as a Function of Recipient Attitude and Source Prestige, J. Abn. Soc. Psychol. 63, 1961, S. 82 - 86; - UPSHAW, H. S.: Own Attitude as an Anchor in Equal Appearing Intervals, J. Abn. Soc. Psychol. 64, 1962, S. 86 - 96; - MANIS, M.: Comment on Upshaw's "Own Attitude as an Anchor in Equal-Appearing Intervals", J. Abn. Soc. Psychol. 68, 1964, S. 689 - 691. - UPSHAW, H. S.: A Linear Alternative to Assimilation-Contrast: A Reply to Manis, J. Abn. Soc. Psychol. 68, 1964, S. 691 - 693; - MANIS, M.: Rejoinder to Upshaw's Reply, J. Abn. Soc. Psychol. 68, 1964, S. 693; - UPSHAW, H. S.: The Effect of Variable Perspectives on Judgment of Opinion Statements for Thurstone Scales: Equal-Appearing Intervals, J. Pers. Soc. Psychol. 2, 1965, S. 60 - 69

4313 Die Technik der nachträglich bestimmten Abstände

Die Technik der gleicherscheinenden Abstände bietet, wie gesagt, keine Möglichkeit zur Prüfung der grundlegenden Annahme, wie weit die Intervalle, in die die einzelnen Items eingestuft werden, auch tatsächlich gleich sind. Sowohl theoretische Überlegungen als auch empirische Untersuchungen zeigen, daß diese Annahme höchst fragwürdig ist: Wenn z.B. für ein Item nach dem Verfahren des Paarvergleichs eine extreme Position auf dem psychologischen Kontinuum von positiv bis negativ errechnet wird, so wird dieses Item bei einer Skalierung nach der Technik der gleicherscheinenden Abstände wahrscheinlich eine weniger extreme Position einnehmen.

Die Technik der nachträglich bestimmten Abstände stellt nun einen Versuch dar, einerseits die Vorteile des Verfahrens der gleicherscheinenden Abstände beizubehalten - daß nämlich eine Versuchsperson nur jeweils ein Urteil für jedes Item abgeben muß und so die Skalierung einer Vielzahl von Items möglich wird - und andererseits die geschilderten Nachteile dieses Verfahrens zu überwinden, d.h. die Errechnung von Punktwerten zu ermöglichen, die Linearität zu den nach der Technik gleicherscheinender Abstände errechneten Punktwerten auch in den Extremen des Kontinuums aufweisen. Die erste Version dieser ursprünglich von THURSTONE entwickelten Technik wurde 1937 von SAFFIR¹⁾ unter der Bezeichnung "method of successive intervals" veröffentlicht. Inzwischen wurde das Verfahren, zum Teil leicht abgewandelt, unter mehreren Benennungen und mit verschiedenen Variationen beschrieben: "method

1) Vgl. SAFFIR, M. A.: A Comparative Study of Scales Constructed by Three Psychophysical Methods, a.a.O.

of absolute scaling",¹⁾ "method of graded dichotomies"²⁾ und "equal discriminability scale".³⁾ In diesem Abschnitt wird nach SCHEUCH die Bezeichnung "Verfahren der nachträglich bestimmten Abstände" verwendet, weil damit sogleich ein wesentlicher Unterschied zur Technik der gleicherscheinen den Abstände herausgestellt wird, handelt es sich doch hier eigentlich nur um "eine andere Berechnungsweise für Urteile von Experten, die mittels der Versuchsanordnung beim Verfahren der gleichen Abstände erhalten wurden ... Skalierungsaufgabe dieser Technik ist die Bestimmung der Abstände zwischen Aussagen, die nach dem zuerst erwähnten Verfahren ausgewählt wurden".⁴⁾

Lie bereits angeführt, ist das Gesetz der kategorialen Urteile als Modifikation des Gesetzes der vergleichenden Urteile die eigentliche Grundlage der Technik der nachträglich bestimmten Abstände. Dieses Gesetz stammt in seiner allgemeinsten Formulierung von TORGERSON⁵⁾. Er geht

- 1) Vgl. GUILFORD, J. P.: The Computation of Psychological Values from Judgments in Absolute Categories, J. Exp. Psychol. 22, 1936, S. 32 - 42
- 2) Vgl. ATTNEAVE, F. A.: A Method of Graded Dichotomies for the Scaling of Judgments, a.a.O.
- 3) Vgl. GARNER, W. R. und H. W. HAKE: The Amount of Information in Absolute Judgments, Psychol. Review 58, 1951, S. 446 - 459
- 4) SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, a.a.O., S. 357
- 5) Vgl. TORGERSON, W. S.: A Law of Categorical Judgment, American Psychologist 9, 1954, S. 483
Vgl. ders.: Theory and Methods of Scaling, a.a.O., S. 235 ff.
Die in diesem Abschnitt verwendete Schreibweise der Formeln findet sich in ganz ähnlicher Form bei UPSHAW, H. S.: Attitude Measurement, in: H. M. BLALOCK Jr. und A. B. BLALOCK (Hrsg.): Methodology in Social Research, a.a.O., S. 86 f. Diese Schreibweise wurde bereits bei der Darstellung des Gesetzes der vergleichenden Urteile verwendet.

dabei davon aus, daß kategoriale Urteile eine Art Paarvergleich zwischen einem zu beurteilenden Item und von im voraus nicht festgelegten Kategoriengrenzen bedeuten. In diesem Falle wird von der Annahme ausgegangen, daß sowohl die Items als auch die Kategoriengrenzen Diskriminationsprozesse hervorrufen, die normal verteilt sind. Auf diese Weise gelangt TORGERSON zu einer Formel für das Gesetz der kategorialen Urteile, die der Formel 4 im Abschnitt über die Technik des Paarvergleichs entspricht.¹⁾ In der neuen Formel wird lediglich der mit dem zweiten Item j verbundene Diskriminationsprozeß S_j durch den mit der Kategoriengrenze j verbundenen Diskriminationsprozeß C_j ersetzt. Für die modalen Diskriminationsprozesse ergibt sich somit:

$$(1) \quad \bar{S}_i - \bar{C}_j = z_{ij} \sqrt{\sigma_i^2 + \sigma_j^2 - 2r_{ij} \sigma_i \sigma_j}$$

Dabei ist: $\bar{S}_i$ = modaler Diskriminationsprozeß des Items i;

$\bar{C}_j$ = modaler Diskriminationsprozeß der Kategoriengrenze j;

z_{ij} = z_{ij} -Wert, der dem empirisch ermittelten Anteil der Urteile: $i > j$ an der Gesamtzahl der Urteile entspricht;

σ_i, σ_j = Standardabweichung von S_i und C_j ;

r_{ij} = Korrelation zwischen S_i und C_j .

Analog zu den Überlegungen bei der Technik des Paarvergleichs gelangt man dann durch vereinfachende Annahmen zu verschiedenen lösbaren Formen der Gleichung (1). An dieser Stelle sei lediglich die Gleichung angeführt, auf der die anschließend erörterte Berechnung der Skalenwerte

1) Vgl. den Abschnitt 4311, S. 104 dieser Arbeit.

für die Items und die Kategoriengrenzen beruht; sie entspricht dem Fall V der THURSTONEschen Modells¹⁾:

$$(2) \quad z_{ij} = \bar{S}_i - \bar{C}_j$$

Die Formulierung dieser Gleichung setzt voraus, daß die Diskriminationsstreuungen für alle Items und für alle Kategoriengrenzen gleich sind. Ebenso müssen alle Korrelationen zwischen den Items und den Kategoriengrenzen gleich sein.

Bei der Konstruktion einer Skala mit Hilfe der Technik nachträglich bestimmter Abstände bekommen die Versuchspersonen dieselben Anweisungen, wie sie bei der Technik gleicherscheinender Abstände gegeben werden können.²⁾ Die grundlegenden Daten für die Berechnung der Skalenwerte der Items sind damit dieselben wie bei der Technik gleicherscheinender Abstände und können auch in ähnlicher Weise wie in der Tabelle 6 des Abschnittes 4312 aufgezeichnet werden.³⁾ Eine solche Tabelle enthält auch in diesem Falle drei Zeilen für jedes Item: Die erste Zeile wird durch eine Häufigkeitsverteilung gebildet, die zeigt, wie oft jedes Item je-

1) Vgl. Formel 8 im Abschnitt 4311, S. 105 dieser Arbeit.

2) Erörterungen der exakten Vorgehensweise für die Berechnung der Skalenwerte der Items und die Kategoriengrenzen bei dieser Technik werden in den meisten einschlägigen Lehrbüchern nicht gegeben. Die Diskussion dieses Abschnittes folgt im wesentlichen der ausgezeichneten Darstellung bei EDWARDS, A. L.: *Techniques of Attitude Scale Construction*, a.a.O., S. 123 - 147. Die Darstellung EDWARDS an dieser Stelle geht teilweise auf einen Aufsatz zurück, den er gemeinsam mit THURSTONE veröffentlicht hat.

Vgl. EDWARDS, A. L. und L. L. THURSTONE: *An Internal Consistency Check for Scale Values Determined by the Method of Successive Intervals*, *Psychometrika* 17, 1952, S. 169 - 180.

Vgl. auch EDWARDS, A. L.: *The Scaling of Stimuli by the Method of Successive Intervals*, *J. Appl. Psychol.* 36, 1952, S. 118 - 122

3) Vgl. Tabelle 6, S. 125 dieser Arbeit.

der der sukzessiven Kategorien von A bis K (1 bis 11) zugeordnet wurde (f). In die zweite Zeile werden die von links nach rechts kumulierten Häufigkeiten eingetragen (cf), und in der dritten Zeile schließlich werden diese kumulativen Häufigkeiten (cf) in kumulative Anteile an der Gesamtzahl der abgegebenen Urteile umgeformt (cp).

Das erste Problem bei einer Skalierung nach diesem Verfahren besteht nun in der Errechnung der Intervallbreiten; denn die, wie zu zeigen versucht wurde, höchst problematische Annahme, daß die Versuchspersonen die Abstände zwischen den einzelnen Kategorien gleich groß wahrnehmen, wird bei dieser Technik aufgegeben. Voraussetzung ist lediglich, daß die Items einer Kategorie j eine positivere Einstellung gegenüber einem bestimmten Einstellungsobjekt ausdrücken als die Items der angrenzenden Kategorie j-1 und eine negativere Einstellung als die Items der angrenzenden Kategorie j + 1. Über die Abstände zwischen den einzelnen Kategorien werden keine Annahmen gemacht: "The mathematical model treats the categories as contiguous segments of the attitude continuum, separated by category boundaries."¹⁾

Zur Berechnung der Intervallbreiten, die insgesamt das psychologische Kontinuum bilden, wird die Verteilung der kumulativen Anteile zugrunde gelegt. Entsprechend den bereits erläuterten Annahmen sind diese kumulativen Anteile für jedes Item normal verteilt. Als Punktwerte eines Items auf dem psychologischen Kontinuum von negativ bis positiv wird der Median oder das arithmetische Mittel aus der Verteilung der kumulativen Anteile berechnet; Voraussetzung für diese Berechnung ist aber zunächst die Abschätzung der Intervall-

1) Green, B. F.: Attitude Measurement, a.a.O., S. 347

breiten. Dabei wird ähnlich wie bei der Technik des Paarvergleichs vorgegangen. Ausgehend von der ersten Tabelle wird eine P-Matrix angefertigt, die nur die kumulativen Anteile der Urteile für die einzelnen Items enthält: Die so gebildete zweite Tabelle besteht aus p_{ij} -Werten, wobei sich das Subskript i auf das Item und das Subskript j auf die obere Grenze der Kategorie j bezieht. Ein bestimmter p_{ij} -Wert zeigt bei dieser Berechnungsweise also an, welcher Anteil der Versuchspersonen das Item i unterhalb der oberen Grenze der Kategorie j eingestuft hat, und $1 - p_{ij}$ ist der Anteil der Versuchspersonen, die das Item i oberhalb der oberen Grenze der Kategorie j eingestuft hat. Anhand der p_{ij} -Werte der P-Matrix kann nun eine Z-Matrix aufgestellt werden, in der die p_{ij} -Werte in z_{ij} -Werte umgeformt werden. Das Subskript i bezieht sich dabei wiederum auf ein bestimmtes Item und das Subskript j auf die obere Grenze der Kategorie j . Als Maß für den Abstand zwischen den angrenzenden Kategorien des psychologischen Kontinuums wird die Differenz zwischen den beiden Eintragungen z_{ij} und $z_{i(j-1)}$ in der Z-Matrix verwendet. Die Formel zur Berechnung der Intervallbreite w_{ij} der Kategorie j , wie sie sich anhand der Urteilsverteilung für Item i berechnen läßt, lautet somit:

$$(3) \quad w_{ij} = z_{ij} - z_{i(j-1)}$$

Anhand dieser Formel können nun die w_{ij} -Werte als Maß für die Intervallbreite für jede der angrenzenden Kategorien berechnet und in eine weitere Tabelle eingetragen werden. Die Eintragungen in den Zeilen dieser Tabelle bilden dann einen Maßstab für die Breite angrenzender Intervalle auf der Grundlage der Verteilung der kumulativen Anteile des

betreffenden Items i ; die Eintragungen in den Spalten der Tabelle sind alle Maßstab für ein und dasselbe Intervall auf der Grundlage der Verteilung der kumulativen Anteile verschiedener zu skalierender Items.

Items	2 - 1	3 - 2	4 - 3		7 - 6	8 - 7
1		.540	.305		.884	.387
2					1.302	1.379
3		.737	.407		.598	.578
.						
.						
14					.835	1.519
(1) Summe	1.599	5.855	4.668		11.039	10.311
(2) n	2	9	10		14	12
(3) $\bar{w}_{.j}$	.800	.651	.467		0.788	.859
(4) $\bar{w}_{.j}$ ku- muliert	.800	1.451	1.918		3.885	4.744

Tabelle 7: Daten zur Berechnung der Intervallbreiten $\bar{w}_{.j}$ ¹⁾

Es wird dann angenommen, daß der beste Maßstab für die Breite der verschiedenen angrenzenden Intervalle das arithmetische Mittel der Eintragungen in den Spalten dieser Tabelle ist. Wie bei der Technik des Paarvergleichs werden auch hier p_{ij} -Werte, die größer als 0,90 oder kleiner als 0,02 sind, zur Berechnung außer acht gelassen; da-

1) Hierbei handelt es sich um die verkürzte Niedergabe des in der Tabelle 5.4 bei EDWARDS enthaltenen Beispiels.
Vgl. EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 123.
Die Tabelle enthält die Differenzen zwischen den z_{ij} -Werten angrenzender Intervalle.

durch wird die Tabelle, wie das Beispiel oben zeigt, unvollständig. Die erste Zeile unter der Tabelle (1) enthält die Spaltensummen. In der zweiten Zeile (2) ist die Zahl der Eintragungen pro Spalte verzeichnet und in der dritten Zeile (3) die arithmetischen Mittel der Eintragungen in den einzelnen Spalten $\bar{w}_{.j}$. Diese werden, wie gesagt, als Maß für die Intervallbreite der betreffenden Kategorien verwendet. Das zweite Intervall hat somit eine Breite von 0.800, das dritte eine Breite von 0.651 usw. Die Verteilung der Intervalle über das gesamte psychologische Kontinuum kann aus den kumulierten $\bar{w}_{.j}$ -Werten der Zeile (4) ersehen werden; dabei wird die Obergrenze des ersten Intervalls, die gleich der Untergrenze des zweiten Intervalls ist, als Ausgangspunkt für die Kumulierung gewählt.

Wenn die Intervallgrenzen in der dargestellten Weise berechnet sind, können die Punktwerte für die einzelnen Items nach der Formel für den Median berechnet werden:

$$(4) \quad S_i = g + \left(\frac{0,50 - \sum p_u}{p_i} \right) \bar{w}_{.j}$$

Dabei bedeutet:

S_i = Punktwert des Items i

g = die Untergrenze des Intervalls, in dem der Median liegt

$\sum p_u$ = die Summe der Anteile unterhalb des Intervalls, in dem der Median liegt

p_i = Anteil innerhalb des Intervalls, in dem der Median liegt

$\bar{w}_{.j}$ = die Breite des Intervalls

Bei der Benutzung des Medians zur Errechnung des Skalenswertes der Items besteht nun eine Schwierigkeit: Wenn die

Items k Kategorien zugeordnet werden, so können anhand dieser Zuordnung lediglich die Intervallbreiten der mittleren $k-2$ Kategorien, nicht aber die Intervallbreiten der beiden Extremkategorien berechnet werden, weil die Untergrenze des ersten Intervalls und die Obergrenze des letzten Intervalls nicht nach der geschilderten Berechnungsweise bestimmbar sind. Das bedeutet, daß die Skalenwerte von Items, die von den Experten zu 50 und mehr Prozent in eines der beiden extremen Intervalle eingestuft werden, nicht ohne zusätzliche Annahmen berechnet werden können. EDWARDS beschreibt eine Möglichkeit, auch in den genannten Fällen Skalenwerte für Items, deren Median in einer der beiden extremen Kategorien liegt, zu berechnen.¹⁾

Wenn ein Item in 50 oder mehr Prozent aller Fälle der ersten Kategorie zugeordnet wird, geht man folgendermaßen vor: Auch bei dieser Berechnungsweise werden in Anlehnung an die Technik des Paarvergleichs nur Eintragungen in Höhe von mindestens 4 % in der ersten Kategorie berücksichtigt, um allzu große Diskrepanzen in den Verhältnissen der p_{ij} - und der z_{ij} -Werte zueinander zu vermeiden.²⁾ Es wird nun angenommen, daß die Hälfte der auf das erste Intervall entfallenden Urteile ($p/2$) einen Schätzwert für den Mittelpunkt dieses Intervalls bildet.³⁾ Die Werte $p/2$

1) Vgl. EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 131 - 135

2) Vgl. dazu die entsprechenden Ausführungen im Abschnitt 4311, S. 113 dieser Arbeit.

3) EDWARDS bemerkt dazu, daß bei den Vorgehensweisen, die das arithmetische Mittel aus der Verteilung der Anteile als Skalenwert für die Items errechnen, eine solche Annahme nicht nötig ist, weil die Vorgehensweise dann nicht von p_{ij} -Werten über 50 % in einer der extremen Kategorien beeinflusst werde. ATTNEAVE und GREEN z.B. verwenden das arithmetische Mittel zur Berechnung der Skalenwerte.

Vgl. ATTNEAVE, F. A.: A Method of Graded Dichotomies for the Scaling of Judgments, a.a.O.; - GREEN, B. F.: Attitude Measurement, a.a.O., S. 347 f.

für jedes Item werden dann anhand einer Normalverteilungstabelle in Z-Werte transformiert. Die Differenz zwischen diesen Z-Werten für $p/2$ und den Z-Werten für p , die die obere Grenze des ersten Intervalls bilden, wird dann als Maß für den Abstand zwischen dem Mittelpunkt des ersten Intervalls bis zur Obergrenze dieses Intervalls - zunächst auf der Grundlage jedes einzelnen Items - benutzt. Wenn für jedes der zu skalierenden Items die Z-Werte für $p/2$ und p bezogen auf das erste Intervall ermittelt sind, wird die durchschnittliche Differenz zwischen diesen Werten als endgültiger Schätzwert für den Abstand zwischen Mittelpunkt und Obergrenze des ersten Intervalls den Berechnungen der Skalenwerte für die Items zugrunde gelegt. Eine solche Berechnungsweise bedeutet - wie EDWARDS betont - eine Ausdehnung des psychologischen Kontinuums nach links und eine Verlegung des Ausgangspunktes von der oberen Grenze des ersten Intervalls, die gleich der Untergrenze des zweiten Intervalls ist, auf den Mittelpunkt des ersten Intervalls. Wird der Mittelpunkt des ersten Intervalls als Ausgangspunkt verwendet, so ist anhand der Formel (4) die Berechnung von Skalenwerten für alle die Items möglich, die von mehr als 50 % der Versuchspersonen in die erste Kategorie eingestuft wurden.

Zur Berechnung des Skalenwertes eines Items, das zu mehr als 50 % der l e t z t e n Kategorie zugeordnet worden ist, wird ähnlich vorgegangen: Grundlage der Berechnung ist der Abstand zwischen der Untergrenze des letzten Intervalls, die gleich der Obergrenze des vorletzten Intervalls ist, und dem Mittelpunkt des letzten Intervalls. Die Werte des kumulativen Anteils der Urteile (cp) für das letzte Intervall sind notwendigerweise 1,00. Um daher den Anteil der Fälle im letzten Intervall zu ermitteln, wird

der c_p -Wert des vorletzten Intervalls von 1,00 abgezogen.¹⁾ Die Zahl der Fälle im letzten Intervall wird nun halbiert ($P/2$) und dem c_p -Wert für die Untergrenze des Intervalls hinzuaddiert; damit ist der c_p -Wert für den Mittelpunkt des letzten Intervalls berechnet. Die c_p -Werte für die Untergrenze und den Mittelpunkt des letzten Intervalls werden dann in der bekannten Weise in Z-Werte transformiert. Die Differenz dieser Z-Werte ist dann, nachdem das arithmetische Mittel aus den Differenzen bei den einzelnen Items berechnet wurde, der Schätzwert für den Abstand zwischen Untergrenze und Mittelpunkt des letzten Intervalls. Mit Hilfe der Formel (4) können nunmehr auch die Punktwerte für alle Items berechnet werden, die von den Versuchspersonen zu mehr als 50 % dem letzten Intervall zugeordnet wurden. Mit der beschriebenen Vorgehensweise ist eine Ausdehnung des psychologischen Kontinuums über die Obergrenze des vorletzten Intervalls hinaus verbunden.

Wenn die Punktwerte für eine Reihe von Items mit Hilfe der Technik der nachträglich bestimmten Abstände berechnet wurden, können die empirischen Daten nach einem ähnlichen Verfahren wie beim Paarvergleich auf interne Konsistenz geprüft werden. Dabei soll kontrolliert werden, wie weit die empirisch gefundenen p_{ij} -Werte mit den p_{ij}' -Werten übereinstimmen, die sich aus den berechneten Punktwerten sowie - im Fall der Technik der nachträglich bestimmten Abstände - aus den Intervallbreiten ableiten lassen.²⁾ Es soll also die empirische Verteilung der

1) In die weitere Berechnung werden dann wiederum nur Werte von mindestens 4 % im letzten Intervall einbezogen.

2) Vgl. EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 37 f. und S. 135 - 138
Vgl. auch S. 110 ff. dieser Arbeit.

p_{ij} -Werte einer aus den berechneten Punktwerten und Intervallbreiten gewonnenen abgeleiteten Verteilung von p_{ij} -Werten gegenübergestellt werden. Ist die Differenz zwischen beiden Verteilungen gering, so besteht Grund zu der Annahme, daß das dieser Skalierungstechnik zugrunde liegende Modell den empirischen Daten adäquat ist.

Zur Durchführung einer solchen Kontrolle stehen bei der Technik nachträglich bestimmter Abstände also die Punktwerte von n skalierten Items und von $k-2$ Intervallbreiten zur Verfügung. Mit Hilfe dieser Parameter läßt sich eine abgeleitete Verteilung¹⁾ von p_{ij} -Werten für die skalierten Items berechnen. Dabei wird zunächst eine Tabelle erstellt, in der links (zur Bezeichnung der Zeilen) die für die Items 1 bis n errechneten Punktwerte untereinander verzeichnet werden. Die Spalten der Tabelle werden mit den errechneten kumulierten Intervallbreiten überschrieben. Jetzt werden die Punktwerte jedes Items von den kumulierten Intervallbreiten subtrahiert und die Ergebnisse in die Tabelle eingetragen. Damit erhalten wir eine abgeleitete Matrix von z_{ij} -Werten.²⁾ Anhand einer Normalverteilungstabelle lassen sich diese abgeleiteten z_{ij} -Werte in die gesuchten abgeleiteten p_{ij} -Werte umformen. Diese p_{ij} -Werte werden in eine weitere Matrix eingetragen. Die Eintragungen in den Zeilen dieser Matrix bilden nun die aus den Punktwerten der Items und den Intervallbreiten abgeleitete Verteilung von p_{ij} -Werten.

-
- 1) Wie bei der Diskussion der Technik des Paarvergleichs spricht EDWARDS auch hier von "theoretical distribution", "theoretical proportion" usw.
Vgl. EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 136 ff.
An dieser Stelle sollen jedoch ebenso wie im Abschnitt 4311 die Ausdrücke "abgeleitete" Verteilung usw. verwendet werden.
 - 2) "These z_{ij} values will be the boundaries of the successive intervals as expressed in terms of normal deviates from the scale values of the statements on the psychological continuum."
EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 137

Evidenz für die Haltbarkeit der Annahmen, die der Technik nachträglich bestimmter Abstände zugrunde liegen, kann durch einen Vergleich zwischen der abgeleiteten Verteilung der p_{ij} '-Werte und der empirisch gefundenen Verteilung der p_{ij} -Werte erbracht werden: Je stärker die beiden Verteilungen übereinstimmen, um so mehr spricht für die Haltbarkeit der Annahmen. Zur Prüfung der Übereinstimmung wird die Differenz zwischen den einzelnen p_{ij} - und p_{ij} '-Werten gebildet. Die absoluten Werte dieser Differenzen werden addiert und durch die Gesamtzahl aller Differenzen dividiert, um so ein Maß für die absolute durchschnittliche Abweichung zu bekommen.¹⁾ Die typischen Werte für dieses Maß liegen etwas über 0,020. Bei EDWARDS und THURSTONE²⁾ ergibt sich ein Wert von 0,025 bei der Skalierung von zehn Items unter Verwendung von neun Kategorien, EDWARDS³⁾ berichtet einen Wert von 0,021 bei der Skalierung von 17 Items und der Verwendung von zehn Kategorien.

Nachdem die Punktwerte für die einzelnen Items berechnet sind, werden zur Verwendung in der endgültigen Skala die Items - wie bei der Technik gleicherscheinender Abstände und beim Paarvergleich - so ausgesucht, daß sie gleichmäßig über das Einstellungskontinuum von negativ bis positiv verteilt sind. Als zweites Auswahlkriterium dient wiederum der Q-Wert, der auf die gleiche Weise wie bei der Technik gleicherscheinender Abstände errechnet wird.⁴⁾ Die Items

1) Selbstverständlich ließen sich auch andere Maße verwenden.

2) Vgl. EDWARDS, A. L. und L. L. THURSTONE: An Internal Consistency Check for Scale Values Determined by the Method of Successive Intervals, a.a.O.

3) Vgl. EDWARDS, A. L.: The Scaling of Stimuli by the Method of Successive Intervals, a.a.O.

4) Vgl. Formel (2) in Abschnitt 4312. 3. 197 dieser Arbeit.

mit den geringsten Q-Werten werden in die Skala aufgenommen. Wie beim Paarvergleich, so kann natürlich auch hier ein anderes Maß für die Differenzen in den Reaktionen der Versuchspersonen, die als Experten eingesetzt werden, Verwendung finden.¹⁾ Am einfachsten bestimmbar sind jedoch die Q-Werte für die einzelnen Items. Die Anweisungen für die Versuchspersonen, deren Einstellung zu einem Einstellungsobjekt mit der endgültigen Skala gemessen werden soll, sind, wie gesagt, dieselben wie bei der Anwendung eines mittels der Technik des Paarvergleiches oder der gleicherscheinenden Abstände konstruierten Instrumentes. Der Punktwert einer Versuchsperson ist der Median oder das arithmetische Mittel der von ihr akzeptierten Behauptungen.

In der Literatur herrscht weitgehende Übereinstimmung darüber, daß mit der Technik der nachträglich bestimmten Abstände eher eine Intervallskala konstruiert werden kann als mit der Technik der gleicherscheinenden Abstände; auf jeden Fall aber läßt sich eine Ordinalskala konstruieren. Diese Annahme wird bestätigt durch die immer wieder festgestellte fast vollständige Linearität der Punktwerte von Items, die nach den Verfahren nachträglich bestimmter Abstände und des

1) Vgl. hierzu z.B. verschiedene Artikel, die alle in Psychometrika erschienen sind:
SAFFIR, M. A.: A Comparative Study of Scales Constructed by Three Psychophysical Methods, a.a.O.; - MOSIER, C. I.: A Modification of the Method of Successive Intervals, Psychometrika 5, 1940, S. 101 - 107; - GULLIKSEN, H.: A Least Squares Solution for Successive Intervals Assuming Unequal Standard Deviations, Psychometrika 19, 1954, S. 117 - 139; - BURROS, R. H.: The Estimation of the Discriminal Dispersion in the Method of Successive Intervals, Psychometrika 20, 1955, S. 299 - 306; - RIMOLDI, H. J. A. und M. HORMAECHÉ: The Law of Comparative Judgment in Successive Intervals and Graphic Rating Scale Methods, Psychometrika 20, 1955, S. 307 - 318; - DIEDERICH, G. W., S. J. HESSICK und L. R. TUCKER: A General Least Squares Solution for Successive Intervals, Psychometrika 22, 1957, S. 159 - 173

Paarvergleiches berechnet sind. Werden die Punktwerte derselben Items nach der Technik gleicherscheinender Abstände bestimmt, so ergibt sich dagegen die bereits geschilderte Verschiebung der Punktwerte zum Mittelpunkt der Skala hin, verbunden mit einer unterschiedlichen Beurteilung der Items je nach der Einstellung der Versuchspersonen zu dem betreffenden Einstellungsobjekt.¹⁾ Die Diskussion dieses Problems soll hier nicht fortgesetzt werden.

Wie bei den beiden anderen bereits erläuterten "Judgment"-Techniken, so ist auch in diesem Falle die Prüfung der Validität und Reliabilität, die eine Skala in der Anwendung aufweist, dem Konstruktionsverfahren selbst extern. Eine solche Prüfung kann mit Hilfe aller oben geschilderten Vorgehensweisen erfolgen.

432 Die Technik der summierten Einschätzungen

Die in den Sozialwissenschaften am häufigsten verwendete Version der Technik der summierten Einschätzungen wurde von LIKERT²⁾ entwickelt und geht auf Verfahren zurück,

1) Vgl. SAFFIR, M. A.: A Comparative Study of Scales Constructed by Three Psychophysical Methods, a.a.O.
Vgl. ATTNEAVE, F.: A Method of Graded Dichotomies for the Scaling of Judgments, a.a.O.
Vgl. EDWARDS, A. L. und L. L. THURSTONE: An Internal Consistency Check for Scale Values Determined by the Method of Successive Intervals, a.a.O.
Vgl. EDWARDS, A. L.: The Scaling of Stimuli by the Method of Successive Intervals, a.a.O.
Vgl. HOVLAND, C. I. und M. SHERIF: Judgmental Phenomena and Scales of Attitude Measurement: Item Displacement in Thurstone Scales, a.a.O.
Vgl. KELLEY, H. H., C. I. HOVLAND, M. SCHWARTZ und R. P. ABELSON: The Influence of Judges' Attitudes in Three Methods of Scaling, a.a.O.

2) Vgl. LIKERT, R.: A Technique for the Measurement of Attitudes, a.a.O.
Die Arbeit wurde mit geringfügigen Änderungen abgedruckt in: MURPHY, G. und R. LIKERT: Public Opinion and the Individual, New York 1937.
Hier werden auch ausführlicher verschiedene Anwendungsmög-

die in ähnlicher Form bereits vorher auf dem Gebiet des "mental testing"¹⁾ angewendet wurden. Es handelt sich hier um eine Art der Einstellungsmessung, bei der versucht wird, auf dem Wege Personen-orientierten Messens zu einer Ordinalskala oder zumindest zu einer teilweise geordneten Skala zu gelangen; nur die Versuchspersonen, nichtaber die Items werden skaliert. An dieser Stelle soll lediglich das von LIKERT entwickelte Verfahren erläutert und kritisiert werden.²⁾

Um mit Hilfe der Technik der summierten Einschätzungen eine Einstellungsskala zu konstruieren, muß zunächst eine große Zahl von Items für eine Probeskala formuliert werden. Das übliche Verhältnis der Items dieser Probeskala zu den für die endgültige Skala benötigten Items beträgt 4 : 1.³⁾ Die Items sollten entweder klar positiv oder klar negativ formuliert sein; eine Verteilung der Items über das gesamte Kontinuum von positiv nach negativ wird also bei dieser Technik nicht angestrebt. Weil hier auf den Ein-

lichkeiten der von LIKERT entwickelten Skalen erörtert. Die bekanntesten sind die Internationalismus-Skala, die Imperialismus-Skala und die Skala, die Einstellungen gegenüber Negern messen soll. Ein neuerlicher Abdruck der Seiten 44-53 des ursprünglichen Aufsatzes von LIKERT findet sich bei FISHBEIN, M. (Hrsg.): Readings in Attitude Theory and Measurement, a.a.O., S. 90 - 95. Vgl. zu den folgenden Ausführungen auch EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 149 ff. Die Arbeit von EDWARDS enthält die wohl ausführlichste Darstellung der Technik der summierten Einschätzungen in der einschlägigen Literatur.

- 1) Vgl. hierzu Fußnote 4, S. 86 dieser Arbeit.
- 2) Es gibt einfachere Verfahren, die sich von dem hier beschriebenen vor allem dadurch unterscheiden, daß lediglich zwei Antwortkategorien (Zustimmung und Ablehnung) vorgegeben werden und auf eine Itemanalyse verzichtet wird, so daß die Auswahl der Items ausschließlich auf der Intuition des jeweiligen Forschers beruht und eine Prüfung auf "innere Konsistenz" der Items einer Skala somit ebenfalls nicht durchgeführt wird.
- 3) Vgl. SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, a.a.O., S. 358

satz von Experten zur Beurteilung der Position von Items in bezug auf die zu messende Einstellung verzichtet wird, brauchen sich die verwendeten Items nicht unbedingt für die Versuchspersonen klar erkennbar auf die zu messende Einstellung zu beziehen. Die einzelnen Items brauchen also nicht unbedingt "content" bzw. "face validity" aufzuweisen.

Jedem Item werden fünf Antwortkategorien zugeordnet, die von "völlige Zustimmung" und "Zustimmung" über "unentschieden" bis zu "Ablehnung" und "völlige Ablehnung" reichen.¹⁾ Handelt es sich um Items, in denen eine positive Einstellung gegenüber dem Einstellungsobjekt zum Ausdruck kommt, so werden den genannten Antwortkategorien die Zahlenwerte 5, 4, 3, 2, 1 oder auch 4, 3, 2, 1, 0 zugeteilt, was wegen der kleineren Zahlen eine rechnerische Vereinfachung bedeutet. Bei Items, die eine negative Einstellung zu dem betreffenden Einstellungsobjekt ausdrücken, wird das Bewertungssystem umgekehrt. Ursprünglich nahm LIKERT die Bewertung der Antwortkategorien eines Items auf der Grundlage einer Normalverteilungstransformation des Anteils der Versuchspersonen vor, die eine bestimmte Antwortkategorie wählten. Diese komplizierte Vorgehensweise des sogenannten "sigma scoring" wurde jedoch aufgegeben, als sich zeigte, daß eine so gefundene Bewertung mit der Bewertung durch die oben beschriebene einfache Zuordnung

1) Natürlich kann die Zahl der Antwortkategorien auch höher oder niedriger sein; ebenso können die einzelnen Kategorien anders als hier wiedergegeben formuliert werden. Vgl. STOSBERG, K.: Die Selbsteinschätzung der Rezipienten und die Kommunikationsdiskrepanz als Determinanten, des Einstellungswandels im Massenkommunikationsprozeß, a.a.O. In dieser Arbeit wird die Selbsteinschätzung der Versuchspersonen mit Hilfe einer LIKERT-Skala gemessen. Dabei wird den Probanden ein neunteiliges Kontinuum vorgelegt, anhand dessen sie ihre Zustimmung oder Ablehnung zu den Items äußern können. Verschiedentlich wird behauptet, daß bereits drei Antwortkategorien ausreichen und durch eine Vermehrung der Kategorienzahl kaum zusätzliche Informationen gewonnen werden können.

ganzer Zahlen zu den Antwortkategorien in Höhe von 0,99 korrelierte.¹⁾ Im Gegensatz zu den bereits beschriebenen "Judgment"-Techniken müssen für eine LIKERT-Skala monotone Items verwendet werden. Bei dieser Form der Items wird - wie bereits ausgeführt - davon ausgegangen, daß der Punktwert, den eine Versuchsperson bei einem Item und bei der Gesamtheit aller Items erreicht, um so höher liegt, je positiver ihre Einstellung gegenüber dem betreffenden Einstellungsobjekt ist.

Eine auf diese Weise entwickelte Probeskala wird zunächst einer Auswahl von Personen aus der Population vorgelegt, für die man die endgültige Skala konstruieren will. Eine solche Vorgehensweise empfiehlt sich im Hinblick auf die anschließend durchzuführende Itemanalyse, bei der die diskriminierendsten Items ermittelt werden sollen: Ein Item, das bei einer bestimmten Population sehr stark diskriminiert, kann sich bei einer anderen als völlig unbrauchbar erweisen. Die Versuchspersonen des Probesamples nehmen zu den einzelnen Items anhand der vorgegebenen Antwortkategorien aufgrund ihrer eigenen Einstellung zu dem betreffenden Einstellungsobjekt Stellung. Die Technik der summierten Einschätzungen unterscheidet sich also auch in diesem Schritt der Skalenkonstruktion von den bereits besprochenen "Judgment"-Verfahren, bei denen die Experten nicht aufgrund ihrer eigenen Einstellungen zu den Items Stellung nehmen, sondern beurteilen, welche Position einem bestimmten Item auf einem Einstellungskontinuum von positiv bis negativ zukommt.

Der Gesamtpunktwert einer Versuchsperson wird dann durch eine Addition der Punktwerte, die für die einzelnen Items erreicht wurden, errechnet. Weil die Stellungnahmen der Versuchspersonen zu den einzelnen Items als einfache

1) Vgl. hierzu EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 149 - 152

"Ratings" angesehen werden können und der Gesamtpunktwert einer Versuchsperson durch Addition der Einzelpunktwerte gewonnen wird, erhielt dieses Verfahren der Skalenkonstruktion den Namen "Technik der summierten Einschätzungen" ("method of summated ratings").¹⁾

In der Summierung von Einzelpunkten liegt die eigentliche Problematik dieser Technik, weil die dazu an sich notwendige Voraussetzung der Additivität nicht angenommen werden kann. Als Rechtfertigung dafür, daß trotz fehlender Voraussetzungen addiert wird, kann lediglich auf die Brauchbarkeit der so gewonnenen Ergebnisse im Sinne eines "it works" hingewiesen werden. Anstelle einer einfachen Addition der Einzelpunktwerte ist auch eine Gewichtung der einzelnen Items möglich, um zu einem Gesamtpunktwert zu gelangen.²⁾ Auch solche Vorgehensweisen beseitigen jedoch nicht die grundsätzlichen Bedenken, die gegen eine gleich wie geartete Summierung von Einzelpunktwerten erhoben werden müssen.

Eine der Techniken der Itemanalyse wird dann angewendet, um aufgrund der Häufigkeitsverteilung der von den Versuchspersonen erreichten Punktwerte diejenigen Items für die endgültige Skala auszuwählen, die am meisten zwischen Versuchspersonen mit hohen und niedrigen Punktwerten diskriminieren. Mit Hilfe einer solchen Technik sollen alle die Items ausgeschieden werden, die sich nicht auf die zu messende Einstellung beziehen. Die nachfolgende Diskussion

1) Der Ausdruck wurde erstmalig von BIRD verwendet:
vgl. BIRD, C.: Social Psychology, New York 1940

2) Vgl. z.B. die Diskussion bei GULLIKSEN, H.: Theory of Mental Tests, a.a.O., Kap. 20, S. 312 - 362

des Modells, das dem Verfahren der summierten Einschätzungen zugrunde gelegt werden kann, gibt hierüber weiteren Aufschluß. An dieser Stelle kann jedoch bereits gesagt werden, daß die aufgrund der Itemanalyse gesammelte Evidenz niemals vollständig ist; denn trotz dieser Prozedur kann ein Item für die endgültige Skala übrigbleiben, das die zu messende Einstellung nicht erfaßt. Das ist dann der Fall, wenn das betreffende Item eine Ursache oder eine Wirkung der zu messenden Einstellung ist und damit hohe Interkorrelationen zustande kommen.¹⁾

Von den bei der Itemanalyse möglichen Vorgehensweisen sollen lediglich drei kurz beschrieben werden:²⁾ Beim ersten Verfahren werden die Versuchspersonen des Probesamples aufgrund ihrer Gesamtpunktwerte auf einer Probekala geordnet, um so die 25 % (oder 20 %) der Versuchspersonen mit den höchsten und den niedrigsten Punktwerten zu ermitteln. Dann wird für jedes Item der durchschnittliche Punktwert in der hohen Gruppe errechnet und dem durchschnittlichen Punktwert in der niedrigen Gruppe gegenübergestellt. Die Diskriminationsfähigkeit eines Items wird dann definiert als Differenz zwischen den durchschnittlichen Punktwerten der hohen und der niedrigen Gruppe.³⁾

1) Vgl. PHILLIPS, B. S.: Social Research, a.a.O., S. 165

2) Zu den Techniken der Itemanalyse allgemein vgl. DAVIS, F. J.: Item Selection Techniques, in: E. F. LINDQUIST (Hrsg.): Educational Measurement, Washington 1951, S. 266 - 328; - ders.: Item Analysis in Relation to Educational and Psychological Testing, Psychol. Bulletin 49, 1952, S. 97 - 119; - SOLOMON, H. (Hrsg.): Studies in Item Analysis and Prediction, Stanford 1961

3) Ein Beispiel anhand der Antisemitismus-Skala von ADORNO u.a. findet sich bei SELTZ, C., M. JAHODA, M. DEUTSCH und S. W. COOK: Research Methods in Social Relations, a.a.O., S. 164 ff.

Die Gleichung (1) zeigt die Berechnungsweise:

$$(1) \quad DF = \bar{X}_H - \bar{X}_N$$

Dabei ist:

DF = Diskriminationsfähigkeit

$\bar{X}_H$ = durchschnittlicher Punktwert eines Items
in der hohen Gruppe

$\bar{X}_N$ = durchschnittlicher Punktwert eines Items in
der niedrigen Gruppe

Hierbei handelt es sich um ein relativ grobes Maß zur Berechnung der Diskriminationsfähigkeit der verwendeten Items; die folgende Gleichung (2) stellt ein feineres Maß zur Berechnung dar:¹⁾

$$(2) \quad DF = \frac{\bar{X}_H - \bar{X}_N}{\sqrt{\frac{S_H^2}{n_H} + \frac{S_N^2}{n_N}}}$$

Außer den bei Formel (1) erläuterten Symbolen werden hier noch folgende verwendet:

S_H^2 = Varianz der Verteilung der Reaktionen zu
einem Item in der hohen Gruppe

S_N^2 = Varianz der Verteilung der Reaktionen zu einem
Item in der niedrigen Gruppe

n_H = Anzahl der Versuchspersonen in der hohen Gruppe

n_N = Anzahl der Versuchspersonen in der niedrigen
Gruppe

1) Es handelt sich hier um eine Form des t-Tests, über dessen Interpretation die meisten Lehrbücher der Statistik Auskunft geben.
Vgl. z.B. BLALOCK, H. M. Jr.: Social Statistics, New York, Toronto und London 1960, S. 144 ff.
Die hier verwendete Schreibweise findet sich bei EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 152

Der Wert von DF stellt ein Maß dafür dar, wie weit ein Item zwischen der hohen und der niedrigen Gruppe differenziert. Als Faustregel kann man nach EDWARDS davon ausgehen, daß ein DF-Wert von 1,75 oder größer ein Item als ausreichend diskriminierend ausweist, vorausgesetzt, daß sich sowohl die hohe als auch die niedrige Gruppe aus mindestens 25 Versuchspersonen zusammensetzt.¹⁾

Ein dritter Weg schließlich zur Bestimmung der Diskriminationsfähigkeit eines Items besteht in der Bildung von Korrelationen zwischen den Einzelpunktwerten für jedes Item und den von den jeweiligen Versuchspersonen erreichten Gesamtpunktwerten. Dabei werden dann die Items mit den höchsten positiven Korrelationen zwischen Einzelpunktwert und Gesamtpunktwert für die endgültige Skala ausgewählt. Obwohl dieses Verfahren den Vorteil hat, daß die Punktwerte sämtlicher Versuchspersonen des Probesamples und nicht nur die der oberen und der unteren Gruppe in die Rechnung eingehen, dürfte es sich hier doch um eine relative aufwendige Vorgehensweise handeln, deren Ergebnisse kaum von den mit den einfacheren Verfahren erzielten Resultaten abweichen werden.²⁾ Da ein wichtiger Vorteil bei der Konstruktion einer LIKERT-Skala vielfach in dem verhältnismäßig geringen Arbeitsaufwand gegenüber anderen Verfahren gesehen wird, sollte man auch bei der Itemanalyse eine möglichst einfache Vorgehensweise wählen.

Die nach der Itemanalyse verbleibenden Items mit der höchsten Diskriminationsfähigkeit werden dann für die endgültige Skala zusammengestellt. Dabei sollte darauf

1) Vgl. EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 153
Vgl. auch die Formel, die GREEN zur Berechnung der "critical ratio" angibt, einer Größe, die dem entspricht, was hier als Diskriminationsfähigkeit bezeichnet wird.

Vgl. GREEN, B. F.: Attitude Measurement, a.a.O., S. 351

2) Vgl. EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 155

geachtet werden, daß etwa eine Hälfte der Items positiv und die andere Hälfte negativ formuliert ist, um der Herausbildung eines bestimmten Reaktionsmusters bei den Versuchspersonen, denen die endgültige Skala vorgelegt wird, vorzubeugen.¹⁾

LIKERT selbst hat kein mathematisches Modell für sein Skalierungsverfahren entwickelt. Es kann jedoch davon ausgegangen werden, daß seine Technik der Skalenkonstruktion auf einer ähnlichen Grundannahme wie die Faktorenanalyse und die Analyse latenter Strukturen aufbaut, auf der Annahme nämlich, daß die Interkorrelationen der Items auf einen einzigen gemeinsamen Faktor zurückzuführen sind, mit dem alle Items gegenseitig verbunden sind.²⁾ Damit ist das Validitätsproblem angeschnitten, und die damit verbundene Frage lautet hier: Aufgrund welcher Annahmen kann man davon ausgehen, daß eine mit der Technik der summierten Einschätzungen konstruierte Skala ein theoretisches Konstrukt so mißt, wie es in der Hypothese, die überprüft werden soll, in seiner konstitutiven Bedeutung formuliert wurde?

Bei unvollständiger Interkorrelation zwischen den Items kann der Punktwert, welcher der Reaktion einer Versuchsperson bei einem einzelnen Item zuerteilt wird, als die gewichtete Summe zweier unabhängiger Faktoren betrachtet werden: Der erste Faktor, der allen Items gemeinsam ist, besteht in der zu messenden Einstellung und der zweite Faktor besteht in einer für jedes Item spezifischen Komponente. Daraus geht hervor, daß mit zunehmender Zahl von

-
- 1) Vgl. die Ergebnisse der Untersuchung von COUCH, A. und K. KENISTON: Yeasayers and Naysayers: Agreeing Response Set as a Personality Variable, J. Abnorm. Soc. Psychol. 60, 1960, S. 151 - 174
- 2) Vgl. zu der folgenden Darstellung GREEN, B. F.: Attitude Measurement, a.a.O., S. 352
GREEN spricht von dem "single common factor model".

Items die lineare Korrelation von Gesamtpunktwert und dem Faktor, der allen Items zugrunde liegt, gegen eins strebt. GREEN gibt an, daß bei 25 Items - LIKERT-Skalen umfassen in der Regel 20 - 30 Items - der typische Wert dieser Korrelation bei 0,9 liege.¹⁾ Der Grund für diese fast vollständige lineare Korrelation besteht darin, daß sich mit der Hinzufügung neuer Items die Wirkung des allen Items zugrunde liegenden Faktors schnell kumuliert. Die Wirkungen der itemspezifischen Faktoren wachsen dagegen nur langsam, weil angenommen wird, daß diese Faktoren unabhängig voneinander variieren und sich in ihrer Wirkung so zumindest teilweise gegenseitig aufheben.

In dem hier kurz beschriebenen Modell wird eine lineare Regression zwischen den Punktwerten, die eine Versuchsperson bei einem Item erreicht, und dem zu messenden gemeinsamen Faktor angenommen. Wenn diese Annahme stimmt, so besteht auch eine lineare Regression zwischen dem Gesamtpunktwert auf einer Skala und dem zu messenden gemeinsamen Faktor. Bei einer unendlichen Zahl von Items besteht damit eine vollständige lineare Korrelation zwischen dem Gesamtpunktwert und dem allen Items gemeinsamen Faktor.²⁾ Aber auch hohe Iteminterkorrelationen können nicht mit Sicherheit als Nachweis dafür angesehen werden, daß mit einer bestimmten nach diesem Verfahren entwickelten Skala auch tatsächlich ein einziger Faktor, d.h. in diesem Falle also eine bestimmte Einstellung, gemessen wird. Denn grund-

1) Vgl. GREEN, B. F.: Attitude Measurement, a.a.O., S. 352

2) Auf die Unzulänglichkeiten dieses Modells, die u.a. in der begrenzten Zahl von Antwortkategorien bei der Vorgehensweise LIKERTs begründet sind, hat LORD hingewiesen.

Vgl. LORD, F. M.: The Relation of Test Score to the Trait Underlying the Test, in: Educ. Psychol. Measurement 13, 1953, S. 517 - 549; abgedruckt in: P. F. LAZARSFELD und N. W. HENRY (Hrsg.): Readings in Mathematical Social Science, Chicago 1966, S. 21 - 53

sätzlich können auch zwei oder mehr in derselben Richtung und derselben Stärke wirksame Faktoren diese Korrelationen herbeiführen.

Mit der Darstellung der dem Verfahren der summierten Einschätzungen zugrunde liegenden Annahmen sollte gezeigt werden, daß nicht endgültig entschieden werden kann, ob eine mit Hilfe dieser Technik konstruierte Skala valide das mißt, was gemessen werden soll. Bei strenger Prüfung der genannten Annahmen läßt sich lediglich Evidenz dafür sammeln, daß der die Punktwerte der Items einer Skala beeinflussende 'singel common factor' mit der zu messenden Variablen in ihrer konstitutiven Bedeutung übereinstimmt. Eine endgültige Entscheidung über die Validität einer Messung ist aber u.a. wegen der schwer abzuschätzenden Wirkungen der itemspezifischen Faktoren nicht möglich.

Bei der Interpretation der Punktwerte im Hinblick auf die Bestimmung von Richtung und Intensität einer Einstellung auf dem Einstellungskontinuum treten gegenüber den bereits beschriebenen "Judgment"-Verfahren einige Besonderheiten auf, die näher zu erläutern sind. Ein bestimmter Punktwert, der nach einem der "Judgment"-Verfahren errechnet wurde, kann ohne Bezugnahme auf eine Gruppe von Versuchspersonen interpretiert werden. Das ist bei den Punktwerten, die mit Hilfe einer nach der Technik der summierten Einschätzungen konstruierten Skala ermittelt wurden, nicht möglich. Hier haben die Punktwerte nur in bezug auf eine jeweils ganz bestimmte Gruppe von Versuchspersonen Bedeutung. Bei einer Skala mit 20 Items, deren Antwortkategorien mit 0 bis 4 bewertet werden, ist der mögliche minimale Gesamtpunktwert einer Versuchsperson 0, der maximale 80. Aufgrund einer gegebenen Verteilung der Gesamtpunktwerte in einer Gruppe lassen sich zwar die extremen Punktwerte um 0 und

um 30 als negative bzw. positive Einstellung gegenüber dem Einstellungsobjekt interpretieren. Bei der Interpretation der Zwischenwerte treten jedoch Schwierigkeiten auf, weil sich kein Punktwert finden läßt, der dem Nullpunkt oder "neutralen" Punkt entspricht, den man z.B. bei der Technik gleicherscheinender Abstände mit einem mittleren oder "neutralen" Item festzulegen glaubt.¹⁾ Bei dem Verfahren der summierten Einschätzungen läßt sich auch keine Evidenz dafür sammeln, daß etwa der mittlere Wert aller möglichen Punktwerte - im obigen Beispiel also ein Wert um 40 - einem solchen Nullpunkt oder "neutralen" Punkt entspricht. Mit dieser Technik läßt sich demnach lediglich eine Ordnung oder Quasiordnung einer Gruppe von Versuchspersonen auf einem Einstellungskontinuum von positiv nach negativ erstellen. Der Gesamtpunktwert, den eine Versuchsperson erreicht, ist also immer nur im Vergleich zu den von anderen Versuchspersonen einer Gruppe erreichten Gesamtpunktwerten interpretierbar.

Ob es somit zweckmäßig ist, eine Skala mit dem Verfahren der summierten Einschätzungen zu konstruieren, hängt von der jeweiligen Untersuchungssituation und speziell von der Form der zu prüfenden Hypothesen ab. Kann man dabei auf die Feststellung eines "neutralen" Punktes verzichten und kommt es nur auf die Ordnung der Versuchspersonen auf einem Einstellungskontinuum an, so läßt sich mit der LIKERT-Skala durchaus arbeiten²⁾ (wenn auch die Voraussetzung für die Addition der Einzelpunktwerte nicht als gegeben angesehen werden kann). Das gilt z.B. auch für die Prüfung der Hypothesen im Rahmen des theoretischen Ansatzes zur Erklärung von Massenkommunikationsprozessen, der in der vorlie-

1) Vgl. EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 157

2) Hilfsweise läßt sich zur Bestimmung einer Richtung in den Einstellungen der Versuchspersonen der mittlere von einer bestimmten Gruppe erreichte Punktwert heranziehen. Eine solche Interpretation kann jedoch immer nur in bezug auf diese eine Gruppe erfolgen.

genden Arbeit erläutert wurde.¹⁾ Besonders in drei Forschungssituationen läßt sich nach EDWARDS die LIKERT-Skala erfolgversprechend einsetzen:²⁾

- a) wenn durchschnittliche Änderungen in den Gesamtpunktwerten als Ergebnis der Einführung eines experimentellen Stimulus verglichen werden sollen,
- b) wenn die durchschnittlichen Gesamtpunktwerte von zwei oder mehr Gruppen verglichen werden sollen und
- c) wenn die Gesamtpunktwerte, die aufgrund einer Skala erreicht werden, mit Gesamtpunktwerten einer anderen Skala korreliert werden sollen.

Um die Vergleichbarkeit der Gesamtpunktwerte zu erleichtern, läßt sich eine "Relativierung" der errechneten Punktwerte aufgrund ihrer Häufigkeitsverteilung in verschiedenen Gruppen von Versuchspersonen durch die Ermittlung sogenannter T-Werte erreichen.³⁾ Dazu wird zunächst der durchschnittliche Gesamtpunkt看wert $\bar{X}$ für eine Gruppe berechnet. Dieser Wert wird als Ausgangspunkt genommen: Höhere Gesamtpunktwerte werden in diesem Fall als positivere, niedrigere als negativere Einstellung - vom Gruppenn-durchschnitt her gesehen - interpretiert. Die Berechnung der T-Werte erfolgt nach der Formel:

$$(3) \quad T = 50 + 10 \left(\frac{X - \bar{X}}{s} \right)$$

1) Vgl. Abschnitt 242 dieser Arbeit.

2) Vgl. EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 157

3) Vgl. ders., a.a.O., S. 158 f.

Dabei ist:

X = Gesamtpunktwert einer bestimmten Versuchsperson

$\bar{X}$ = arithmetisches Mittel der Verteilung der Gesamtpunktwerte in einer Gruppe

s = Standardabweichung der Verteilung der Gesamtpunktwerte in einer Gruppe

Die Umformung der ursprünglichen Gesamtpunktwerte in T-Werte, wie sie bei vielen psychologischen Tests üblich ist, führt zu einer neuen Verteilung, deren Mittelwert 50 und deren Standardabweichung 10 ist. Damit wird eine Interpretation der Gesamtpunktwerte unabhängig von unterschiedlichen Mittelwerten und Standardabweichungen bei den verschiedensten Skalen möglich. Es muß noch einmal betont werden, daß alle diese Operationen an sich ein Skalenniveau voraussetzen, das mit der Technik der summierten Einschätzungen nicht erreicht wird.

Die Frage der Reliabilität von Skalen, die nach den LIKERT-Verfahren konstruiert sind, wird meist erörtert, indem man einen Vergleich mit den "Judgment"-Techniken - insbesondere dem Verfahren der gleicherscheinenden Abstände - durchführt.¹⁾ Dabei handelt es sich um teilweise empirische Un-

1) Vgl. EDWARDS, A. L. und K. C. KENNEY: A Comparison of the Thurstone and Likert Techniques of Attitude Scale Construction, a.a.O.

Vgl. LIKERT, R., S. ROSLOW und G. MURPHY: A Simple and Reliable Method of Scoring the Thurstone Attitude Scales, in: J. Soc. Psychol. 5, 1934, S. 228 - 238

Vgl. EYSENCK, H. J. und S. CROWN: An Experimental Study in Opinion-Attitude Methodology, a.a.O.

Vgl. FERGUSON, L. W.: A Study of the Likert Technique of Attitude Scale Construction, in: J. Soc. Psychol. 13, 1941, S. 51 - 57

Eine zusammenfassende Darstellung befindet sich bei EDWARDS, A. L.: Techniques of Attitude Measurement, a.a.O., S. 159 - 169

tersuchungen, die jedoch die grundlegende Unterscheidung außer acht lassen, daß bei den "Judgment"-Verfahren die Skalierung nichtmonotoner Items vorgenommen wird, während bei der Technik der summierten Einschätzungen monotone Items verwendet werden.¹⁾ Darum werden entsprechende Reliabilitätsvergleiche problematisch, wenn dieselben Items - wie das in einer Reihe der zitierten Untersuchungen geschehen ist - einmal nach dem LIKERT-Verfahren und einmal nach den "Judgment"-Verfahren skaliert werden. Ohne auf die Einzelheiten dabei näher einzugehen, läßt sich ganz allgemein sagen, daß die Reliabilität von Skalen, die mit dem Verfahren der summierten Einschätzungen konstruiert sind, der Reliabilität der nach dem Verfahren der gleicherscheinernden Abstände konstruierten Skalen nicht nachsteht.²⁾ Sehr unterschiedliche Ergebnisse wurden dagegen in dieser Beziehung bei einem Vergleich zwischen LIKERT-Verfahren und Skalogramm-Analyse erzielt.³⁾

Zusammenfassend läßt sich zum Verfahren der summierten Einschätzungen sagen: Es handelt sich hier um eine Technik Personen-orientierten Messens, mit der eine Ordnung oder zumindest eine Quasiordnung der Versuchspersonen in bezug auf die zu messende Einstellung erreicht werden kann. Versuche zur Bestimmung gleicher Intervalle oder zur Kon-

1) Vgl. GREEN, B. F.: Attitude Measurement, a.a.O., S. 365. Diese Tatsache verdeutlicht das Schaubild 1 bei FERGUSON, L. W.: A Study of the Likert Technique of Attitude Scale Construction, a.a.O., S. 56, abgedruckt bei EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 164
Vgl. auch die entsprechenden Ausführungen im Abschnitt 434 dieser Arbeit.

2) Zur Bestimmung der Reliabilität einer LIKERT-Skala wird meist das Halbierungsverfahren angewendet. Bei Vergleichen zwischen LIKERT- und "Judgment"-Verfahren handelt es sich um eine Art, die dem Verfahren der äquivalenten Formen ähnlich ist, wobei jedoch die Einzelskalen trotz der damit verbundenen Problematik zum Teil dieselben Items enthalten.

3) Vgl. CLARK, K. E. und P. H. KRIEDT: An Application of Guttman's New Scaling Techniques to an Attitude

struktion eines Nullpunktes werden nicht unternommen. Was die Bestimmung von Richtung und Intensität einer Einstellung betrifft, so ist eine Interpretation der Punktwerte jeweils nur unter Bezugnahme auf eine genau definierte Gruppe von Versuchspersonen möglich. Für die Validität einer Messung läßt sich anhand des zugrunde liegenden Modells Evidenz sammeln. Die erzielten Reliabilitätsgrade sind denen anderer Meßtechniken vergleichbar.

433 Die Technik der Skalogramm-Analyse

Die Technik der Skalogrammanalyse brachte gegenüber den Ansätzen von THURSTONE und LIKERT eine neue Betrachtungsweise auf dem Gebiet der Einstellungsmessung. Sie wurde in ihrer ursprünglichen Form von GUTTMAN¹⁾ erarbeitet und in einer Fülle von Abhandlungen diskutiert, die sich zum Teil - ausgehend von dem grundlegenden Modell - mit speziellen Fragen der auftretenden Probleme beschäftigen. Manche Autoren behaupten, daß es sich bei der Skalogramm-Analyse nicht eigentlich um eine Skalierungstechnik handle, sondern lediglich um ein Verfahren zur Prüfung der Frage, ob eine Reihe von Items den Anforderungen eines bestimmten Modells - nämlich des von GUTTMAN formulierten - genügt. SCHEUCH drückt das in ähnlicher Form aus: "Die Vorgehensweise bei der Skalogramm-Analyse besteht im Prinzip lediglich aus einer Ober-

Questionnaire, Educ. Psychol. Measmt. 8, 1948, S. 215-223
Vgl. KRIEDT, P. H. und K. E. CLARK: "Item Analysis" versus "Scale Analysis", J. Appl. Psychol. 33, 1949, S. 114-121

- 1) Die erste Arbeit, in der GUTTMAN seine neuen Ideen über die Skalenkonstruktion vortrug, war:
GUTTMAN, L.: The Quantification of a Class of Attributes: A Theory and Method of Scale Construction, a.a.O.
Es folgte eine Reihe von weiteren Arbeiten, z.B.: GUTTMAN, L.: A Basis for Scaling Qualitative Data, Amer. Soc. Rev. 9, 1944, S. 139 - 150; - ders.: The Cornell Technique for Scale and Intensity Analysis, Educ. Psychol. Measmt. 7, 1947, S. 247 - 279
Schließlich findet sich die ausführliche Diskussion der Skalogramm-Analyse von GUTTMAN und SUCHMAN in:
STOUFFER, S. A. u.a.: Measurement and Prediction, Princeton, N.J. 1950, Kapitel 2 - 9, S. 46 - 361

prüfung, ob eine beobachtete Folge von Reaktionen den Bedingungen eines besonders stringenten Modells entspricht. ... So gesehen ist die Skalogramm-Analyse weniger ein Skalierungsverfahren als eine Art Leistungstest, den ein Satz von Reaktionen zu bestehen hat."¹⁾

Mit Hilfe dieser Technik wird versucht, "einen Satz qualitativer Daten zu skalieren, die eine offenbare (!) Beziehung zu einem einzelnen gemeinsamen Faktor besitzen; es handelt sich damit um einen Spezialfall einer 'Faktorenanalyse' für qualitative Daten."²⁾

Bei der Skalogramm-Analyse wird von folgender Annahme ausgegangen: "it must be possible to order the items such that, ideally, persons who answer a given question favorably all have higher ranks than persons who answer the same question unfavorably"³⁾. In GUTTMANs eigenen Worten lautet das so: "We shall call a set of items of common content a scale if a person with a higher rank than another person is just as high or higher on every item than the other person."⁴⁾ Dieser Gedanke lag be-

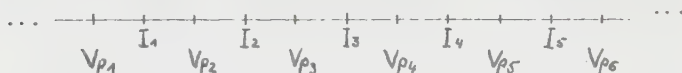
Vgl. auch RILEY WHITE, M., J. W. RILEY Jr. und J. TOBY: Sociological Studies in Scale Analysis, New Brunswick, N. J., 1954

- 1) SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, a.a.O., S. 361
- 2) ders., a.a.O., S. 360
- 3) STOUFFER, S. A.: An Overview of the Contributions to Scaling and Scale Theory, in: S. A. STOUFFER u.a.: Measurement and Prediction, a.a.O., S. 5
- 4) GUTTMAN, L.: The Basis for Scalogram Analysis, in: S. A. STOUFFER u.a.: Measurement and Prediction, a.a.O., S. 62

reits der Entwicklung der berühmten "social distance scale" von BOGARDUS zugrunde, nur wurde er hier nicht zu einem allgemeinen Skalierungsmodell ausgebaut.¹⁾

Bei der Skalogramm-Analyse handelt es sich um reaktions-orientiertes Messen, das heißt, sowohl Items - in diesem Fall monotone²⁾ - als auch Versuchspersonen werden in eine Rangfolge gebracht. Das bedeutet: "the systematic variability of responses of subjects to stimuli is ascribed to variability both in the subjects and in the stimuli."³⁾ Der Idealfall einer Skala nimmt dann in der graphischen Darstellung folgende Form an:⁴⁾

Zunehmende Werte der zu messenden Eigenschaft



Graphische Darstellung einer mit Hilfe der Technik der Skalogramm-Analyse konstruierten Skala, auf der fünf Items (I_1 bis I_5) und sechs Versuchspersonen (V_{P_1} bis V_{P_6}) platziert sind.

Angestrebt wird die Messung auf Ordinalskalenniveau; über die Abstände zwischen den Items auf dem psychologischen

- 1) Vgl. BOGARDUS, E. S.: Measuring Social Distance, J. Appl. Sociol. 9, 1925, S. 299 - 308, abgedruckt in: FISHBEIN, M. (Hrsg.): Readings in Attitude Theory and Measurement, a.a.O., S. 71 - 76
- 2) Inzwischen wurden auch Versuche unternommen, Skalogramm-Analysen bei nichtmonotonen Items durchzuführen. Vgl. MOSTELLER, F.: A Theory of Scalogram Analysis Using Noncumulative Types of Items, Report Nr. 9, Laboratory of Social Relations, Harvard University, 1949 Vgl. MOHLWILL, J.: The Measurement of Scalability for Non-Cumulative Items, Educ. Psychol. Measmt. 23, 1963
- 3) TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 298 Vgl. auch Abschnitt 423 dieser Arbeit.
- 4) Die Darstellung stammt von UPSHAW, H. S.: Attitude Measurement, a.a.O., S. 99, Figur 3.1

Kontinuum und über die Abstände zwischen den in einer Rangfolge geordneten Versuchspersonen wird in der Regel nichts ausgesagt. Wenn eine Reihe von Items den von GUTTMAN gestellten Anforderungen genügt, so bilden diese Items eine eindimensionale Skala.

Werden die Items den Versuchspersonen zur Annahme oder Ablehnung vorgelegt und genügen sie den Anforderungen des Modells, so ist die einzig mögliche Kombination der Antworten die in der Tabelle 8 aufgezeigte:

Vpn ¹⁾	Antwort- schemata	Items a b c d e	Punktwerte der Vpn	Rangplätze der Vpn in bezug auf die zu messende Einstellung
1	I	1 1 1 1 1	5	1
2	II	0 1 1 1 1	4	2
3	III	0 0 1 1 1	3	3
4	IV	0 0 0 1 1	2	4
5	V	0 0 0 0 1	1	5
6	VI	0 0 0 0 0	0	6

Tabelle 8: Einzig mögliche Folge von Antwortschemata, die dem Modell der Skalogramm-Analyse genügt

Eine solche Folge von Antwortschemata (response patterns) kommt folgendermaßen zustande: Eine Reihe von Items (im Beispiel fünf) wird formuliert und einer Gruppe von Versuchspersonen (im Beispiel sechs) vorgelegt. Es wird unterstellt, daß diese Versuchspersonen eine unterschiedliche Einstellung zu dem Einstellungsobjekt haben, auf das sich die Items beziehen. Ihre Position auf dem psychologischen Kontinuum ist jedoch ebenso unbekannt wie die Posi-

1) Vpn steht für Versuchspersonen

tion der formulierten Items. Nun wird gleichsam jede Versuchsperson mit jedem Item verglichen, wobei die Reaktionen der Versuchspersonen, wenn nur zwei Antwortkategorien bestehen, mit 1 oder mit 0 gewichtet werden. Die Gewichtung erfolgt so, daß die 1 jeweils derjenigen Reaktion zugeordnet wird, die ein Mehr der zu messenden Eigenschaft ausdrückt. Für jede Versuchsperson läßt sich dann ein Punktwert als Summe der Gewichte für ihre Reaktionen auf die Items errechnen.¹⁾ Nur unter der Annahme also,²⁾ daß die verwendeten Items auf einer einzigen Dimension angeordnet werden können und die Reaktionen der Versuchspersonen nur durch ihre Position auf dieser Dimension bestimmt werden, ist die einzig mögliche Folge von Antwortschemata die in Tabelle 8 wiedergegebene. Folgen von Antwortschemata, die dem Modell entsprechen, werden als "skalierbare Typen", alle übrigen als "nicht-skalierbare Typen" bezeichnet.³⁾

Die Kombination von Antworten, die die Bildung einer Skala ermöglicht, läßt sich anhand dreier Kriterien beschreiben:

- a) "Pbn, welche ein gegebenes Item positiv beantworten, haben im Skalogramm einen höheren Rangplatz als Pbn, welche das Item negativ beantworten,
- b) wenn von zwei Pbn i und j der Pbn j einen höheren Rangplatz im Skalogramm hat, dann hat er alle Items, welche i positiv beantwortete, ebenfalls positiv beantwortet und darüber hinaus noch so viele Items, als die Rangplatzdifferenz ausmacht,
- c) aus dem Rangplatz eines Pbn kann man bestimmen, welche Items er positiv beantwortet hat."⁴⁾

Dabei sind a) und b) immer gegeben, wenn c) gegeben ist.

-
- 1) Wegen dieser Art der Errechnung von Punktwerten wird eine so konstruierte Skala auch kumulative Skala genannt.
 - 2) SCHEUCH spricht von einer "Art umgekehrter Nullhypothese"; vgl. SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, a.a.O., S. 361
 - 3) Vgl. hierzu weiter unten in diesem Abschnitt.
 - 4) SIXTL, F.: Meßmethoden in der Psychologie, a.a.O., S. 400 f. Pbn steht für Probanden.

Nun läßt sich im voraus natürlich nicht bestimmen, ob eine Reihe von Items die Anforderungen des Modells - auf einer einzigen Dimension skalierbar zu sein - erfüllt oder nicht. Die Prüfung dieser Frage ist die Aufgabe der Skalogramm-Analyse. Wenn diese Analyse das Antwortschema der Tabelle 8 ergibt, "then these results would offer strong evidence that the statements do fall along a single dimension. Thus, if the responses of subjects to the statements were in accord with our theoretical model of a unidimensional scale of statements, we would have confidence in interpreting scores of subjects based upon the statements as also falling along the same unidimensional continuum."¹⁾

Es fragt sich jedoch, ob es nicht möglich ist, bereits vor der Anwendung einer der Techniken der Skalogramm-Analyse, eine Reihe von Items zu formulieren, von der mit einiger Sicherheit angenommen werden kann, daß sie skalierbar ist. Die Ausführungen GUTTMANS zu diesem Problem sind ausgesprochen dürftig. Er sagt lediglich, daß die Auswahl einer kleinen Zahl von Items aus der möglichen großen Zahl aller Items, die ein "Bedeutungskollektiv" ("universe of content") repräsentieren, aufgrund von Intuition und Erfahrung erfolgen solle,²⁾ wobei nach GUTTMAN die Items von dem Versuchsleiter bewußt konstruiert und nicht nach dem Zufallsprinzip aus der möglichen Zahl aller Items ausgewählt werden: REMMERS gibt die Überlegungen GUTTMANS zu diesem Punkt folgendermaßen wieder: "We shall think of these items as a sample from a

1) EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 176

2) Vgl. GUTTMAN, L.: Questions and Answers about Scale Analysis, Research Branch, Information and Education Division, Army Service Forces, Report D-2, 1945; zitiert bei EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 177

hypothetically infinite universe of items of similar content. It is important to notice that this 'sampling' process is highly subjective, and by no means a random one, so that usual sampling theory is of no use to us in interpreting the behavior of our sample of items."¹⁾ Die Bemerkungen GUTTMANs über das Bedeutungskollektiv selbst sind ebenso recht verschwommen: "The universe consists of all the attributes that define the concept ... it consists of all the attributes of interest to the investigation which have a common content, so that they are classified under a single heading which indicates that content."²⁾ In der herkömmlichen Terminologie am klarsten ist wohl die Interpretation bei SCHEUCH, nach der mit Bedeutungskollektiv "der Sinnzusammenhang gemeint (ist), für den ein konkreter Satz von Indikatoren als operationale Definition verstanden werden kann"³⁾.

Es hat sich gezeigt, daß die Chancen, eine Reihe skalierbarer Items zu erhalten, größer ist, je spezifischer ein Bedeutungskollektiv formuliert wurde. Bei dem Versuch einer Skalierung wird in der Regel mit relativ wenigen Items (höchstens zwölf) gearbeitet, was nicht zuletzt darauf zurückzuführen ist, daß mit der Zahl der Items die Zahl nicht-skalierbarer Typen überproportional steigt. Diese Items sollten, wie FESTINGER meint, in großem Aus-

-
- 1) REMNERS, H. H.: Introduction to Opinion and Attitude Measurement, New York 1954, S. 104
 - 2) GUTTMAN, L.: The Basis for Scalogram-Analysis, a.a.O., S. 80
Vgl. hierzu auch die Erörterung bei TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 332
 - 3) SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, a.a.O., S. 361
Zur Diskussion des damit angeschnittenen Validitätsproblems vgl. weiter unten in diesem Abschnitt.

maß immer wieder dasselbe ausdrücken ("rephrasings of the same thing"¹⁾), das heißt, sie sollten in ihrem Inhalt möglichst homogen sein.²⁾ Die Berücksichtigung dieser Hinweise bedeutet natürlich keine Gewähr dafür, daß auch tatsächlich eine Reihe skalierbarer Items formuliert wird. Als sicherstes Mittel zur Erreichung dieses Zieles schlagen deshalb EDWARDS und KILPATRICK eine Form der Itemanalyse vor, die der Skalogramm-Analyse voranzugehen hat.³⁾

Wenn die Items formuliert und den Versuchspersonen als Alternativfragen vorgelegt worden sind, werden die Reaktionen in eine Tabelle, das "Skalogramm", eingetragen, wobei die Gewichtung der Reaktionen in der bereits beschriebenen Weise erfolgt.⁴⁾ Jede Versuchsperson bekommt im Skalogramm eine Zeile, jedes Item eine Spalte. Die Versuchspersonen werden von vornherein nach der zu erwartenden Rangfolge ihrer Gesamtpunktwerte, die auf die bereits geschilderte Art errechnet werden, geordnet. Ebenso werden die Items nach der zu erwartenden Rangfolge auf dem psychologischen Kontinuum geordnet. Aufgrund der Eintragungen in diesem Skalogramm wird nun versucht, die Skalierbarkeit des Beobachtungsmaterials zu prüfen. Dafür steht eine Reihe von Techniken zur Verfügung:

- a) Das Verfahren der kleinsten Quadrate. Diese Vorgehensweise war von GUTTMAN ursprünglich vorgeschlagen worden;⁵⁾ es zeigte sich jedoch, daß sie im Falle der Skalogramm-

1) FESTINGER, L.: The Treatment of Qualitative Data by "Scale Analysis", Psychol. Bulletin 44, 1947, S.149-161

2) Vgl. LOEVINGER, J.: The Technique of Homogenous Tests Compared with Some Aspects of "Scale Analysis" and Factor Analysis, Psychol. Bulletin 45, 1948, S. 507 - 529

3) Vgl. EDWARDS, A. L. und F. P. KILPATRICK: Scale Analysis and the Measurement of Social Attitudes, a.a.O.
Vgl. hierzu auch den folgenden Abschnitt dieser Arbeit über die Diskriminationstechnik.

4) Das Skalogramm hat die Form der Tabelle 8.

5) Vgl. GUTTMAN, L.: The Quantification of a Class of Attributes: A Theory and Method of Scale Construction, a.a.O.

Analyse mit einem unnötigen Arbeitsaufwand verbunden war,¹⁾ denn weniger aufwendige Techniken führten ebenfalls zu brauchbaren Ergebnissen.

- b) Die von GOODENOUGH entwickelte Technik der Häufigkeitsverteilung,²⁾
- c) die von GUTTMAN selbst erarbeitete Cornell Technik,³⁾
- d) mechanische Vorrichtungen wie z.B. das Skalogramm-Brett,⁴⁾
- e) Techniken, bei denen Lochkartenmaschinen eingesetzt werden.⁵⁾

In der vorliegenden Arbeit sollen lediglich die Cornell-Technik und die Technik der Häufigkeitsverteilung als die am häufigsten eingesetzten Verfahren kurz erläutert werden, da sie die grundlegenden Ideen und Probleme der Vorgehensweise bei der Skalogramm-Analyse ausreichend verdeutlichen.

Bei der Verwendung der Cornell Technik zur Prüfung der Skalierbarkeit einer Reihe von Items wird zunächst eine Tabelle erstellt, die - in Abwandlung von der ursprüng-

-
- 1) Vgl. SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, a.a.O., S. 352
 - 2) Vgl. GOODENOUGH, W. H.: A Technique for Scale Analysis, Educ. Psychol. Measmt. 4, 1944, S. 179 - 190
 - 3) Vgl. GUTTMAN, L.: The Cornell Technique for Scale and Intensity Analysis, a.a.O.
 - 4) Vgl. SUCHMAN, E. A.: The Scalogram board Technique, in: S. A. STOUFFER u.a.: Measurement and Prediction, a.a.O., S. 91 - 121
 - 5) Vgl. FORD, R. N.: A Rapid Scoring Procedure for Scaling Attitude Questions, Public Opinion Quarterly 14, 1950, S. 507 - 532
Vgl. KAHN, L. H. und A. J. BODINE: Guttman Scale Analysis by Means of IBM Equipment, Educ. Psychol. Measmt. 11, 1951, S. 288 - 314
Vgl. AKE, J. N.: A Rapid Machine Procedure for Determining Scalability of any Number of Questions, Public Opinion Quarterly 26, 1962, S. 121 - 125

lichen Form - für jede Antwortkategorie jedes Items eine Spalte und für jede Versuchsperson eine Zeile enthält. Ansonsten wird die Anordnung der zuerst beschriebenen Tabelle beibehalten. Wenn GUTTMAN 100 Versuchspersonen und zwölf Items als angemessen für einen Skalierbarkeitstest ansieht, würde die Tabelle somit bei dichotomischen Antwortkategorien 24 Spalten und 100 Zeilen enthalten. Der Einfachheit halber wird in der Tabelle 9 auf der folgenden Seite das konstruierte Beispiel von EDWARDS¹⁾ wiedergegeben, weil es in der Literatur das anschaulichste über diese Technik sein dürfte.

Wenn die 20 Items in diesem Beispiel eine Skala bilden, müssen sie den Anforderungen des Modells genügen. Die erste Anforderung besteht darin, daß "from a person's rank alone we can reproduce his response to each of the items in a simple fashion"²⁾. Das ist jedoch vollständig nur in einem kaum zu erreichenden Idealfall möglich; deswegen ist es notwendig, einen Maßstab für bestimmte Grade der Reproduzierbarkeit zu entwickeln. Ein solcher Maßstab kann im konkreten Fall dann verwendet werden, um über Skalierbarkeit oder Nicht-Skalierbarkeit entscheiden zu können. Zu diesem Zweck werden zunächst sogenannte Wendepunkte ("cutting points") ermittelt: Das sind die Punkte in einer Rangordnung von Versuchspersonen, an denen ihre überwiegenden Reaktionen von einer Antwortkategorie zur anderen wechseln.³⁾ Damit überhaupt solche Wendepunkte errechnet

1) Tabelle 9 wurde leicht geändert und ergänzt von EDWARDS übernommen.

Vgl. EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 183, Tabelle 7.3

2) GUTTMAN, L.: The Cornell Technique for Scale and Intensity Analysis, a.a.O., S. 249

3) Diese Wendepunkte sind in Tabelle 9 durch waagerechte Striche markiert.

Vpn	Items								Punkt- wert	e ₂
	1		2		3		4			
	1	0	1	0	1	0	1	0		
1	x		x		x		x		4	0
2	x			x	x		x		3	0
3	x		x			x	x		3	2
4	x		x			x	x		3	2
5	x			x	x		x		3	0
6	x			x	x		x		3	0
7	x			x	x		x		3	0
8	x			x	x		x		3	0
9	x			x		x	x		2	0
10		x	x			x	x		2	2
11	x			x		x	x		2	0
12	x			x		x	x		2	0
13		x	x			x	x		2	2
14		x		x	x		x		2	2
15		x	x			x		x	1	2
16		x		x		x	x		1	0
17	x			x		x		x	1	2
18		x		x	x			x	1	2
19		x		x		x	x		1	0
20		x		x		x		x	0	0
(1) f	12	8	6	14	8	12	16	4	Σe ₂ = 16	
(2) p und q	0,6	0,4	0,3	0,7	0,4	0,6	0,8	0,2		
(3) e ₁	1	1	3	1	2	2	2	0	Σe ₁ = 12	

Tabelle 9: Die Cornell Technik und die Technik der Häufigkeitsverteilung nach GOODENOUGH für 4 Items und 20 Versuchspersonen¹⁾

1) Für Tabelle 9 gilt:

x = Antworten der Versuchspersonen

f = Zahl der Antworten pro Antwortkategorie

p und q = relative Häufigkeiten der Antworten pro Antwortkategorie

e₁ = Fehler nach der Cornell-Technik berechnet

e₂ = Fehler nach der Technik von GOODENOUGH berechnet

werden können, müssen bei der Skalogramm-Analyse monotone Items verwendet werden, bei welchen im Idealfall die Wahrscheinlichkeit einer positiven Reaktion bis zu einem bestimmten Punkt auf dem psychologischen Kontinuum, dem Wendepunkt, gleich null ist, um von da an auf eins zu springen.¹⁾

GUTTMAN stellt zwei Regeln für die Ermittlung der Wendepunkte auf: Sie sollen so gelegt werden, daß

- a) die Zahl der Fehler (vgl. weiter unten) ein Minimum erreicht und
- b) in keiner Kategorie mehr Fehler als Nicht-Fehler auftreten.²⁾

Wenn die Antworten der Versuchspersonen vollständig aus ihrem Punktwert reproduzierbar sind und die verwendeten Items dem Modell damit völlig entsprechen, müssen alle Antworten oberhalb eines Wendepunktes in die eine Kategorie und alle Antworten unterhalb dieses Wendepunktes in die andere Kategorie fallen. Diejenigen Antworten, die aus diesem Schema herausfallen, werden als Fehler behandelt.

Ist die Auswertung der Daten so weit fortgeschritten, kann ein Reproduktionskoeffizient ("coefficient of reproducibility") berechnet werden:

$$\text{Reproduktionskoeffizient RK} = 1 - \frac{\text{Zahl der Fehler}}{\text{Zahl der Items} \cdot \text{Zahl der Vpn}}$$

1) Diese Form der Reaktionskurve entspricht in etwa der Kurve 3 in Figur 2, S. 82 dieser Arbeit. Bei der Kurve 3 liegen jedoch die Wahrscheinlichkeiten einer positiven Reaktion etwas über null bzw. etwas unter eins, was der praktischen Forschungssituation eher entspricht als der aufgezeigte Idealfall.

2) Vgl. GUTTMAN, L.: The Cornell Technique for Scale and Intensity Analysis, a.a.O., S. 258 und 261

In diesem Ausdruck wird also die Gesamtzahl der Fehler zur Gesamtzahl der Antworten ins Verhältnis gesetzt und von 1 abgezogen. Für das Beispiel der Tabelle 9 ergibt sich somit:

$$RK = 1 - \frac{12}{4 \cdot 20} = 0,85$$

Werte von 0,85 und mehr werden im allgemeinen als ausreichend angesehen, um die Skalierbarkeit der Items anzunehmen. Der Reproduktionskoeffizient soll angeben, mit welcher Genauigkeit (in Prozent) die Reaktionen auf die einzelnen Items aufgrund der von den Versuchspersonen erreichten Punktwerte rekonstruiert werden können.¹⁾ Die Daten der Tabelle 9 zeigen jedoch, daß eine Rekonstruktion der individuellen Reaktionen mit dieser Genauigkeit nicht möglich ist. Der für die angegebenen Daten errechnete Reproduktionskoeffizient zeigt deswegen eine zu hohe Genauigkeit (in Prozent) an, weil die Regel befolgt wurde, nach der bei der Ermittlung der Wendepunkte auf Minimierung der Fehler zu achten ist.²⁾

Die von GOODENOUGH³⁾ entwickelte Technik zur Durchführung einer Skalogramm-Analyse bietet dagegen die Möglichkeit, den Reproduktionskoeffizienten so zu errechnen, daß er die Genauigkeit, mit der die individuellen Reaktionen auf die Items anhand der Punktwerte der Versuchspersonen rekonstruierbar sind, exakt wiedergibt.⁴⁾ Bei dieser Technik

1) Die Parallele zu den Verfahren der Prüfung auf innere Konsistenz der Daten, wie sie im Zusammenhang mit der Technik des Paarvergleiches und der Technik der nachträglich bestimmten Abstände erläutert wurden, liegt auf der Hand; vgl. die Abschnitte 4311 und 4313 dieser Arbeit.

2) Vgl. EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 184

3) Vgl. GOODENOUGH, W. H.: A Technique for Scale Analysis, a.a.O.

4) Vgl. EDWARDS, A. L.: On Guttman's Scale Analysis, Educ. Psychol. Measmt. 8, 1948, S. 313 - 318

die hier nicht ausführlich erläutert werden soll, werden zunächst die relativen Häufigkeiten der mit 1 gewichteten Reaktionen pro Item als P-Werte ermittelt; der Anteil der mit 0 gewichteten Reaktionen pro Item ist dann $q = 1-p$.¹⁾ Mit Hilfe eines speziellen graphischen Verfahrens werden dann aufgrund der so festgestellten P- und Q-Werte die Antwortschemata für jeden einzelnen Punktwert und damit auch für jede einzelne Versuchsperson, die in einer Tabelle nach diesen Punktwerten eingeordnet sind, vorausgesagt. Im Anschluß daran werden die vorausgesagten Antwortschemata mit den tatsächlich beobachteten verglichen. Jede Abweichung einer beobachteten Antwort von einer vorausgesagten wird als Fehler gezählt. Es handelt sich hier also um eine andere Art der Fehlerfeststellung als bei der Cornell Technik. Sie erlaubt, wie gesagt, die Berechnung eines Reproduktionskoeffizienten, der exakt angibt, mit welcher Genauigkeit die individuellen Reaktionen aus den Punktwerten reproduziert werden können.²⁾

Nun ist aber der Nachweis eines Reproduktionskoeffizienten von 0,85 und höher zwar eine notwendige, aber keine ausreichende Bedingung, um die Skalierbarkeit der Items nachzuweisen. Der Grund dafür ist einfach: Die Reproduzierbarkeit der Reaktionen auf ein Item kann niemals

1) Vgl. die zweite Zeile unter der Tabelle 9.

2) Für die Daten der Tabelle 9 ergibt sich nach dem von GOODENOUGH entwickelten Verfahren ein Reproduktionskoeffizient von:

$$RK = 1 - \frac{16}{4 \cdot 25} = 0,80$$

Vgl. EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 188

Zu den einzelnen Verfahren zur Bestimmung der Reproduzierbarkeit und der damit zusammenhängenden Probleme vgl. auch WHITE, B. W. und E. SALTZ: Measurement of Reproducibility, Psychol. Bulletin 54, 1957, S. 81 - 99

niedriger sein als die von den Versuchspersonen am häufigsten gewählte Antwortkategorie, die modale Kategorie. Wenn die Häufigkeiten, mit denen die einzelnen Antwortkategorien gewählt wurden, bekannt sind, läßt sich der kleinste für eine Reihe von Items mögliche Reproduktionskoeffizient folgendermaßen bestimmen: Die Werte für die bei den einzelnen Items am häufigsten gewählten Antwortkategorien - ausgedrückt als relative Häufigkeiten der modalen Kategorien - werden addiert¹⁾ und durch die Zahl der zu skalierenden Items dividiert. Auf diese Weise erhält man den Wert der **minimalen marginalen Reproduzierbarkeit** ("minimal marginal reproducibility"²⁾). Für das Beispiel der Tabelle 9 ergibt sich somit:

$$\text{Minimale marginale Reproduzierbarkeit} = \frac{0,6 + 0,7 + 0,6 + 0,8}{4} = \frac{2,7}{4} = 0,675$$

Aus diesen Überlegungen geht hervor, daß die Qualität eines Reproduktionskoeffizienten als Indikator für die Skalierbarkeit einer Reihe von Items vom Wert der minimalen marginalen Reproduzierbarkeit abhängig ist. Beträgt der Wert dieser minimalen marginalen Reproduzierbarkeit z.B. 0,9 und wird ein Reproduktionskoeffizient von z.B. 0,95 errechnet, so ist 0,95 eine Scheingröße, die darum eine ungewöhnlich hohe Reproduzierbarkeit vortäuscht, weil sie auf die hohe minimale marginale Reproduzierbarkeit zurückzuführen ist. Bei einer Skalogramm-Analyse sollte demnach außer dem Wert des Reproduktionskoeffizienten auch immer der Wert der minimalen marginalen Reproduzierbarkeit

1) Vgl. Zeile 2 unter der Tabelle 9.

2) Vgl. GUTTMAN, L.: On Festinger's Evaluation of Scale Analysis, Psychol. Bull. 44, 1947, S. 461 - 465
Vgl. EDWARDS, A. L. und F. P. KILPATRICK: Scale Analysis and the Measurement of Social Attitudes, a.a.O.

errechnet werden. Denn nur die Differenz zwischen diesen beiden Werten, die eine bestimmte durch Konvention zu fixierende Größe nicht unterschreiten sollte, gibt Auskunft über die Skalierbarkeit einer Reihe von Items. Eine gleichmäßige Verteilung der Antworten der Versuchspersonen auf die vorgegebenen Kategorien läßt die Wahrscheinlichkeit eines trügerisch hohen Reproduktionskoeffizienten sinken. Andererseits sind natürlich Items mit ungleicher Antwortverteilung nötig, wenn eine Skalierung überhaupt erfolgreich durchgeführt werden soll.¹⁾

Aber auch wenn bei einer Reihe von Items der Reproduktionskoeffizient ausreichend hoch und der Wert der minimalen marginalen Reproduzierbarkeit im Vergleich dazu niedrig ist, besteht noch keine ausreichende Evidenz dafür, daß die verwendeten Items eine Skala bilden, die den von GUTTMAN formulierten Anforderungen genügt. Denn sobald der Reproduktionskoeffizient nicht den Wert von 1,00 erreicht, was in der Praxis immer der Fall sein wird, liegen Fehler im Sinne des Modells vor. Diese Fehler können zufällig sein oder auch durch die Wirksamkeit einer ursprünglich nicht in Betracht gezogenen Variable verursacht werden. Ob das eine oder das andere oder beides der Fall ist, kann nur durch eine Analyse der nicht-skalierbaren Typen (Antwortschemata) im oben beschriebenen Sinne festgestellt werden.

1) Auf die Fragen der Skalogramm-Analyse bei mehr als zwei Antwortkategorien soll hier nicht näher eingegangen werden, zumal dabei keine grundsätzlich neuen Probleme entstehen. Lediglich im Zusammenhang mit der Berechnung des Reproduktionskoeffizienten und seiner Beurteilung können einige Schwierigkeiten entstehen, die eventuell eine Zusammenfassung der Antwortkategorien angebracht erscheinen lassen. Vgl. die zusammenfassende Darstellung bei EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 188 - 191

Die Gesamtzahl möglicher Typen hängt von der Zahl der zu skalierenden Items und von der Zahl der Antwortkategorien ab. Bei nur zwei Antwortkategorien beträgt die Zahl der möglichen Typen 2^n , wenn n die Anzahl der Items ist. Bei fünf Items ergeben sich somit 32 Typen; davon sind aber nur $n + 1$ Typen mit dem Modell vereinbar, das der Technik der Skalogramm-Analyse zugrunde liegt. Das sind bei fünf Items also sechs von insgesamt 32 möglichen Typen. Weil mit zunehmender Zahl von Items die Zahl nicht-skalierbarer Typen schneller steigt als die Zahl skalierbarer Typen, werden bei der Skalogramm-Analyse im Verhältnis zu anderen Techniken immer relativ wenige Items verwendet. Das ist ein Nachteil des Verfahrens, der sich auf die Validität der Messung negativ auswirken kann, wenn man bedenkt, daß z.B. bei der Technik der summierten Einschätzungen davon ausgegangen wird, daß erst 20 bis 25 Items eine Messung von "ausreichender" Validität ermöglichen.

Wenn nun bei der Analyse der Antwortschemata, die nicht dem Modell entsprechen, der nicht-skalierbaren Typen also, ein Typ oder mehrere Typen gefunden werden, die häufig auftreten, so sieht GUTTMAN darin ein Zeichen für die Existenz mindestens einer weiteren Variablen, welche die zu skalierenden Items ebenso repräsentieren wie die ursprünglich angenommene Variable. In einem solchen Fall empfiehlt GUTTMAN die Aufspaltung des Bedeutungskollektivs, wie es vor Beginn der Skalogramm-Analyse definiert wurde, in zwei oder mehr "Unter-Kollektive", um so neue Skalen entwickeln zu können, mit deren Hilfe jede der Variablen getrennt gemessen werden kann.¹⁾ Es ist dann natürlich möglich, Interkorrelationen zwischen den von einer Versuchsperson erreichten Punktwerten bei den verschiedenen Skalen zu errechnen.

1) Vgl. GUTTMAN, L.: On Festinger's Evaluation of Scale Analysis, a.a.O.

Aus der Literatur geht allerdings nicht hervor, mit welcher Häufigkeit bestimmte nicht-skalierbare Typen auftreten müssen, damit man annehmen kann, daß das ursprüngliche Bedeutungskollektiv aufgespalten werden muß, weil außer dem anfangs angenommenen Faktor noch ein weiterer den Skalierungsvorgang beeinflußt.

Werden bei einem Reproduktionskoeffizienten von weniger als 0,90 bzw. 0,85 keine häufig auftretenden nicht-skalierbaren Typen ermittelt, so nennt man die zu skalierende Reihe von Items eine Quasi-Skala.¹⁾ Es kann dann angenommen werden, daß die Skalierung von einer "durchschlagenden" Variablen und mehreren weniger wirksamen Variablen beeinflußt wird und so eine relativ geringe Reproduzierbarkeit feststellbar ist.²⁾ Diese Ausführungen zeigen, daß auch ein Reproduktionskoeffizient von 0,90 und mehr keine ausreichende Gewähr dafür bietet, daß eine Reihe von Items die Anforderungen des Modells erfüllt. Außer dem Reproduktionskoeffizienten müssen verschiedene weitere Kriterien zur Entscheidung dieser Frage herangezogen werden.³⁾

-
- 1) Vgl. SUCHMAN, E. A.: The Utility of Scalogram Analysis, in: S. A. STOFFER u.a.: Measurement and Prediction, a.a.O., S. 159 - 163
 - 2) Vgl. hierzu auch die Aufsätze von SAGI und SCHUESSLER, in welchen Signifikanztests zur Prüfung der Frage erarbeitet werden, ob die empirisch vorgefundene Verteilung der Antworten durch Zufall zustande gekommen ist. Vgl. SAGI, P. C.: A Statistical Test for the Significance of the Coefficient of Reproducibility, Psychometrika 24, 1959, S. 19 - 27
Vgl. SCHUESSLER, K. F.: A Note on Statistical Significance of Scalogram, Sociometry 24, 1961, S. 312 - 328
 - 3) Es sind verschiedene Techniken zur Verbesserung kumulativer Skalen entwickelt worden, die in dieser Arbeit nur z. T. erörtert werden können. Die bekanntesten sind die Diskriminationstechnik von EDWARDS und KILPATRICK, wie sie im nachfolgenden Abschnitt dieser Arbeit behandelt wird, die H-Technik, die von STOFFER u.a. entwickelt wurde, sowie die auf EDWARDS zurückgehende W-Technik. Vgl. STOFFER, S. A., E. F. BORGATTA, D. G. HAYS und A. F. HENRY: A Technique for Improving Cumulative Scales, Public Opinion Quarterly 16, 1952, S. 273 - 291

Allgemein besteht Übereinstimmung darüber, daß es mit Hilfe der Technik der Skalogramm-Analyse gelingt, eine - allerdings nur relativ kleine - Zahl von Items sowie die eingesetzten Versuchspersonen gemäß der zu messenden Einstellung in eine Rangfolge zu bringen, was einer Messung auf Ordinalskalenniveau entspricht. Versuche, eine Intervallskala zu konstruieren oder einen Nullpunkt zu errichten, werden nicht unternommen. "GUTTMAN's scaling theory constitutes a strict adherence to the logical structure of an ordinal scale ... The result of a GUTTMAN analysis on data which satisfy the necessary conditions is an ordinal scale with the stimuli and the response patterns of the individuals simply ordered."¹⁾ Im Vergleich mit der Technik der gleicherscheinenden Abstände und der Technik der summierten Einschätzungen stellt PHILLIPS sogar fest, daß die von GUTTMAN entwickelte Vorgehensweise die einzige ist, die definitive Kriterien zur Entscheidung der Frage aufstellt, wieweit ein bestimmtes Meßinstrument die einer Ordinalskala zugrunde liegenden Postulate erfüllt.²⁾ Wenn es darüber hinaus möglich ist, die Intensität einer Einstellung unabhängig zu erfassen, kann ein Nullpunkt z.B. als der Punktwert definiert werden, bei dem die Intensität ein Minimum erreicht.³⁾ Bei der Technik der Skalogramm-

Vgl. EDWARDS, A. L.: A Technique for Increasing the Reproducibility of Cumulative Attitude Scales, a.a.O.

- 1) COOMBS, C. H.: Theory and Methods of Social Measurement, a.a.O., S. 526
- 2) Vgl. PHILLIPS, B. S.: Social Research, a.a.O., S. 189 und 203
- 3) Die Einzelheiten finden sich bei GUTTMAN, L.: The Cornell Technique for Scale and Intensity Analysis, a.a.O.; - GUTTMAN, L. und E. A. SUCHMAN: Intensity and a Zero Point for Attitude Analysis, Amer. Soc. Rev. 12, 1947, S. 57 - 67; - SUCHMAN, E. A.: The Intensity Component in Attitude and Opinion Research, in: S. A. STOUFFER u.a.: Measurement and Prediction, a.a.O., S. 213 - 276; - CAMPBELL, E. Q.: Scale and Intensity Analysis in the Study of Attitude Change, Public Opinion Quarterly 26, 1962, S. 227 - 235

Analyse werden, wie gesagt, besondere Anstrengungen unternommen, die Eindimensionalität einer Skala sicherzustellen;¹⁾ ein Maß dafür ist der Reproduktionskoeffizient.

Es wird behauptet, daß die Skalogramm-Analyse "usually yields scales that are reliable and valid according to the usual estimates of these attributes"²⁾. Das mag in bezug auf die Reliabilität einer Messung zutreffen.³⁾ Als Prüfungstechniken bei der Anwendung einer Skala kommen vor allem das Wiederholungsverfahren und das Verfahren äquivalenter Formen in Frage, während das Halbierungsverfahren und seine Varianten u.a. wegen der geringen Zahl der bei einer Skalogramm-Analyse benutzten Items möglicherweise weniger geeignet ist. Die Berechtigung der Bemerkung von SHAW und WRIGHT dürfte jedoch, was die Validität der Messung betrifft, sehr stark von der Definition dessen abhängen, was GUTTMAN das "Bedeutungskollektiv" nennt. Hierzu liegen, wie bereits gesagt, nur recht vage Ausführungen vor. Jedenfalls handelt es sich bei dem Bedeutungskollektiv um einen Begriff, der ex ante in seiner konstitutiven Bedeutung festgelegt wird, das heißt also vor der Formulierung der Items und der Durchführung der Skalogramm-Analyse. Die Items bilden dann die operationale Definition des Begriffes. Ist nun die Skalogramm-Analyse abgewickelt und festgestellt worden, daß die verwendeten Items eine Skala im Sinne GUTTMANs bilden, so ist damit noch nicht gesagt,

1) Zu den Ausweitungen des Modells auf den Bereich der multidimensionalen Skalierung vgl. die zusammenfassende Darstellung bei TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 345 - 359 und die dort angegebene Literatur.

2) SHAW, M. E. und J. M. WRIGHT: Scales for the Measurement of Attitudes, a.a.O., S. 26

3) Vgl. die ausführliche Diskussion der Reliabilitätsprobleme bei der Skalogramm-Analyse von GUTTMAN, L.: Problems of Reliability, in: S. A. STOUFFER u.a.: Measurement and Prediction, a.a.O., S. 277 - 311

daß der Begriff, auf den sich die Items beziehen, in seiner konstitutiven und in seiner operationalen Bedeutung übereinstimmen muß. Auch bei dieser Technik kann erst durch die üblichen Verfahren der Validitätskontrolle entschieden werden, ob es gerechtfertigt erscheint, eine mit Hilfe der Skalogramm-Analyse konstruierte Skala als valides Instrument zur Messung der Eigenschaft, die gemessen werden soll, anzusehen. Die Frage, ob eine Reihe von Items die Bedingungen des Modells der Skalogramm-Analyse erfüllt, hängt mit dem Problem der Validität, wie es in dieser Arbeit verstanden wird, in folgender Weise zusammen: Wenn eine Reihe von Items die Bedingungen des Modells nicht erfüllt, so kann gefolgert werden, daß die Reaktionen auf diese Items noch von mindestens einem weiteren Faktor außer dem zu messenden beeinflußt werden. Das bedeutet, daß das Instrument als invalide in bezug auf die zu messende Eigenschaft abgelehnt werden kann. Genügt jedoch eine Reihe von Items den Anforderungen des Modells, so zeigt das noch nicht, daß sie als operationale Definition eines Begriffes mit der konstitutiven Bedeutung dieses Begriffes übereinstimmt. Wie gesagt, muß eine Prüfung dieser Frage mittels der geschilderten Verfahren erfolgen. Auch in diesem Falle zeigt sich also wieder, daß eine endgültige Validierung eines Meßinstrumentes nicht möglich ist.

REMMERS faßt die Kritik an der Skalogramm-Analyse unter Stöhwort "Mangel an methodologischer Präzision" zusammen und zählt im einzelnen als Mängel auf: "the absence of objective, quantified criteria for establishing cutting points, testing for random scatter of error, testing for degree of 'spuriousness' in obtained reproducibility coefficients due to high modal frequencies, etc."¹⁾ Damit

1) REMMERS, H. H.: Introduction to Opinion and Attitude Measurement, a.a.O., S. 128

bleibt die Entscheidung, ob die empirischen Daten den Anforderungen des Modells entsprechen, zum großen Teil der "Intuition"¹⁾ des jeweiligen Forschers überlassen. Da es sich um ein deterministisches und nicht um ein Wahrscheinlichkeitsmodell handelt, ist zudem die Unterscheidung zwischen zufälligen und systematischen Fehlern nur schwer durchführbar.²⁾ REIMERS fügt die Kritik an der subjektiven Weise der Itemformulierung an: "it is improper to infer from the scalability of a particular set of items that the universe of content which the investigator had in mind is also scalable; it may only be some relatively narrow subuniverse that has been shown to be scalable."³⁾ Das trifft - in einer anderen Terminologie - genau das, was bereits zum Problem der Validitätsprüfung gesagt wurde.

Hingewiesen sei noch auf die Möglichkeit von zwei weiteren Mängeln. Es ist denkbar, daß sich eine Reihe von Items bei einer Gruppe von Versuchspersonen als skalierbar erweist, bei einer anderen dagegen nicht.⁴⁾ Deshalb ist es angebracht, die Skalierbarkeit von Items - ähnlich der Technik der summierten Einschätzungen - jeweils bei der Gruppe von Versuchspersonen zu prüfen, deren Einstellungen dann auch mit Hilfe der endgültigen Skala gemessen werden sollen.⁵⁾ Ebenso kann die Skalierbarkeit einer Reihe von Items zu einem bestimmten Zeitpunkt möglich sein, zu einem anderen dagegen nicht.

1) Vgl. UPSHAW, H. S.: Attitude Measurement, a.a.O., S. 103

2) Vgl. SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, a.a.O., S. 363
Vgl. GREEN, B. F.: Attitude Measurement, a.a.O., S. 357

3) REIMERS, H. H.: Introduction to Opinion and Attitude Measurement, a.a.O., S. 128

4) Ein illustratives Beispiel bietet die Untersuchung von: HARDING, J. und R. HOGREFE: Attitude of White Department Store Employees toward Negro Co-Workers, J. Social Issues 8, 1952, S. 18 - 28

5) Vgl. die Bemerkungen von SHAW, M. E. und J. M. WRIGHT: Scales for the Measurement of Attitudes, a.a.O., S. 26

434 Die Diskriminationstechnik

Bei der Diskriminationstechnik werden Elemente der Verfahren gleicherscheinender Abstände bzw. der nachträglich bestimmten Abstände sowie der Technik der summierten Einschätzungen kombiniert, um eine Skala zu konstruieren, die den von GUTTMAN für die Skalogramm-Analyse aufgestellten Anforderungen entspricht.¹⁾ Bei dieser von EDWARDS und KILPATRICK²⁾ entwickelten Technik geht es also darum, eine kumulative Skala zu konstruieren. Einer der Hauptkritikpunkte an der Skalogramm-Analyse war der, daß bei dieser Technik kaum ein Hinweis für die Formulierung und Auswahl der zu skalierenden Items gegeben wird. Das eigentliche Anliegen der Diskriminationstechnik besteht nun darin, die Auswahl der Items für die Skalogramm-Analyse so vorzubereiten, daß mit größerer Wahrscheinlichkeit, als das ohne Anwendung dieser Technik möglich wäre, die Skalierbarkeit der Items angenommen werden kann: "It would appear as if the important problem is to be able to select a set of statements, in advance of applying the technique of scalogram analysis, that one may with some hope count upon to meet the requirements of a Guttman scale."³⁾ Dieses Problem versuchen EDWARDS und KILPATRICK, wie gesagt, durch Kombination bestimmter Elemente der Techniken gleicherschei-

-
- 1) Vgl. SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, a.a.O., S. 359
 - 2) Vgl. EDWARDS, A. L.: A Critique of "Neutral Items" in Attitude Scales Constructed by the Method of Equal-Appearing Intervals, Psychol. Review 53, 1946, S. 159 - 169
Vgl. EDWARDS, A. L. und F. P. KILPATRICK: A Technique for the Construction of Attitude Scales, a.a.O.
Vgl. dies.: Scale Analysis and the Measurement of Social Attitudes, a.a.O.
Die Darstellung dieses Kapitels stützt sich im wesentlichen auf den Abschnitt: "The Scale-Discrimination Technique", in: EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 201 - 219. Diesem Abschnitt seines Buches hat EDWARDS die maßgebenden Artikel zur Diskriminationstechnik zugrunde gelegt.
 - 3) EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 202 f.

nender Abstände bzw. nachträglich bestimmter Abstände und der summierten Einschätzungen, die der eigentlichen Skalogramm-Analyse vorgeschaltet werden, zu lösen.

Den Ausgangspunkt für ihren Vorschlag der Skalenkonstruktion bilden Untersuchungen über die Eigenart von Items bei verschiedenen Skalierungstechniken. Das Ergebnis dieser Untersuchungen kann nach REMMERS¹⁾ folgendermaßen zusammengefaßt werden: EDWARDS und KILPATRICK beobachteten zwei empirische Beziehungen:

- a) Die Wendepunkte und die Antwortverteilung von Items sind mit ihren Skalenwerten, wenn diese nach dem Verfahren der gleicherscheinenenden Abstände berechnet werden, in der Weise verbunden, daß die Items an den Extrempunkten einer Skala eine hohe modale Antworthäufigkeit (P- oder Q-Werte) aufweisen, während die Verteilung der Antworten bei neutralen Items eher gleichmäßig ist ($p = 0,5$ und $q = 0,5$).
- b) Die Reproduzierbarkeit der Reaktionen auf ein Item aus den Punktwerten der Versuchspersonen ist mit der Diskriminationsfähigkeit dieses Items, die mit einer der Techniken der Itemanalyse ermittelt wird, in der Weise verbunden, daß die Reproduzierbarkeit um so größer ist, je stärker das Item zwischen extremen Gruppen diskriminiert.

Weiter wurde für neutrale Items eine geringere Reproduzierbarkeit als für Items an den Extrempunkten einer Skala festgestellt. Das kann darauf zurückzuführen sein, daß bei neutralen Items die Wahrscheinlichkeit der Zustimmung von Personen auf entgegengesetzten Enden eines Einstellungskontinuums mehr oder weniger gleich ist. Außerdem scheint nach den Untersuchungen von EDWARDS und KILPATRICK das von THURSTONE vorgeschlagene Kriterium für die Auswahl der

1) Vgl. REMMERS, H. H.: Introduction to Opinion and Attitude Measurement, a.a.O., S. 129

Items (Q-Werte) in keiner Beziehung zu der Diskriminationsfähigkeit eines Items zu stehen.

Aufgrund dieser Untersuchungsergebnisse schlagen EDWARDS und KILPATRICK folgende Vorgehensweise für die Konstruktion einer Skala vor: Zunächst wird - dem Verfahren bei der Technik gleicherscheinender Abstände folgend - eine große Zahl von Items, die sich auf die zu messende Variable beziehen, formuliert. Diese Items werden dann einer Gruppe von Versuchspersonen mit der Anweisung vorgelegt, sie auf einem Kontinuum, das in neun oder elf Kategorien eingeteilt ist und von negativ bis positiv reicht, einzustufen. Dabei kann jedes der bei den Techniken der gleicherscheinenden Abstände oder der nachträglich bestimmten Abstände gebräuchlichen Verfahren verwendet werden,¹⁾ um die Urteile der Versuchspersonen einzuholen. Aufgrund dieser Urteile werden dann die Skalenwerte für die einzelnen Items berechnet: Weil die Grundannahme der Technik gleicherscheinender Abstände - daß nämlich gleicherscheinende Abstände auch gleich groß sind - für die Diskriminationstechnik nicht von so großer Bedeutung ist,²⁾ können die Skalenwerte der Items wie bei der Technik der gleicherscheinenden Abstände errechnet werden, um so den bei der Technik nachträglich bestimmter Abstände notwendig größeren Arbeitsaufwand zu vermeiden. Im Anschluß daran wird ein Q-Wert für jedes Item sowie der Median der Q-Werte aller Items errechnet. Die 50 % der Items, die über dem Median der Q-Werte aller Items liegen, werden danach ausgesondert, um auf diese Weise diejenigen Items auszuschneiden, welche die größte Streuung der Urteile auf dem psychologischen Kontinuum aufweisen.

1) Vgl. die Abschnitte 4312 und 4313 dieser Arbeit.

2) Vgl. EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., S. 210

Der Rest der Items (50 %) mit den niedrigeren Q-Werten wird dann zu einer Skala zusammengestellt, wobei jedes Item mit mehreren Antwortkategorien versehen wird. EDWARDS und KILPATRICK benutzten sechs Kategorien: "Völlige Zustimmung", "Zustimmung", "mäßige Zustimmung", "mäßige Ablehnung", "Ablehnung" und "völlige Ablehnung". Diese Aufteilung kommt der später nötigen Dichotomisierung der Antwortkategorien entgegen; es können jedoch auch weniger Kategorien als hier beschrieben vorgegeben werden. Diese Items werden nun einer neuen Gruppe von 200 bis 300 Versuchspersonen vorgelegt. Die Versuchspersonen werden aufgefordert, ihre Zustimmung oder Ablehnung gegenüber den in den Items enthaltenen Behauptungen anhand der vorgegebenen Antwortkategorien zum Ausdruck zu bringen. Die Antworten der Versuchspersonen werden dann so gewichtet, wie es beim Verfahren der summierten Einschätzungen üblich ist: in der Untersuchung von EDWARDS und KILPATRICK mit 0 bis 5. Die Gewichtung erfolgt so, daß die Antwortkategorie, die die positivere Einstellung gegenüber dem betreffenden Einstellungsobjekt zum Ausdruck bringt, den größeren Zahlenwert erhält. Wie jedes Item gewichtet werden muß, läßt sich aus den zuvor berechneten Skalenwerten ersehen. Für jede Versuchsperson wird dann ein Gesamtpunktwert als Summe der Einzelpunktwerte errechnet.

Aufgrund der so ermittelten Daten wird anschließend jedes Item einer Analyse unterzogen. Dazu werden die 27 % der Versuchspersonen mit den höchsten und die 27 % der Versuchspersonen mit den niedrigsten Gesamtpunktwerten ausgewählt, um die Verteilung der Antworten für diese beiden Gruppen getrennt erfassen zu können, wie das am Beispiel eines Items in Tabelle 10 geschehen ist.

Antwortkategorien	Gewichtung	niedrige Gruppe f	hohe Gruppe f
Völlige Zustimmung	5	3	38
Zustimmung	4	5	42
mäßige Zustimmung	3	8	15
mäßige Ablehnung	2	26	2
Ablehnung	1	36	2
völlige Ablehnung	0	22	1
Vpn		100	100

Tabelle 10: Verteilung der Antworten bei einem Item für die hohe und die niedrige Gruppe mit je 100 Versuchspersonen¹⁾

EDWARDS und KILPATRICK schlagen nun vor, an diesem Punkt des Skalierungsvorganges eine Dichotomisierung der Antwortkategorien der Items vorzunehmen, um den später zu erwartenden Schwierigkeiten einer Dichotomisierung während der eigentlichen Skalogramm-Analyse zu entgehen. Diese Dichotomisierung erfolgt in der Weise, daß eine horizontale Linie zwischen den Antwortkategorien der Tabelle 10 gezogen wird, und zwar so, daß die Zahl der Versuchspersonen aus der niedrigen Gruppe oberhalb der Linie und die Zahl der Versuchspersonen aus der hohen Gruppe unterhalb der Linie ein Minimum erreichen. Im Beispiel der Tabelle 10 liegt die Linie, welche die Versuchspersonen in der dargestellten Art trennt, zwischen der mit 3 und der mit 2 gewichteten Antwortkategorie. Das bedeutet: Wenn das Item der

1) Die Tabellen 10 und 11 dieses Abschnittes sind entnommen von EDWARDS, A. L.: Techniques of Attitude Scale Construction, a.a.O., Tabelle 8.1, S. 212 und Tabelle 8.2, S. 214

Tabelle 10 einer Skalogramm-Analyse unterzogen wird, so ordnet man den bisher mit 5, 4, und 3 gewichteten Kategorien die Zahl 1 und den bisher mit 2, 1 und 0 gewichteten Kategorien die Zahl 0 zu. Dasselbe gilt entsprechend für alle anderen Items. Eine allgemeine Darstellung ist in Tabelle 11 enthalten:

Antwortkategorien	niedrige Gruppe	hohe Gruppe	insgesamt
	a	b	a + b
	c	d	c + d
insgesamt	a + c	b + d	

Tabelle 11: Allgemeine Darstellung für die Dichotomisierung von Antwortkategorien, wenn mehr als zwei Kategorien vorgegeben waren

Unabhängig von der Zahl der Antwortkategorien soll die horizontale Linie in der Tabelle 11 so gelegt werden, daß die Häufigkeiten a + d ein Minimum bilden.

Der nächste Schritt bei dieser Art der Skalenkonstruktion besteht in der Messung der Diskriminationsfähigkeit der Items. EDWARDS und KILPATRICK verwenden wegen der Einfachheit der Berechnung den phi-Koeffizienten¹⁾, der sich nach erfolgter Dichotomisierung der Antwortkategorien anhand der Eintragungen in Tabelle 11 ermitteln läßt. Die Formel dafür lautet:

$$r\phi = \frac{bc - ad}{\sqrt{(a + b)(b + d)(a + c)(c + d)}}$$

Für die Werte der Tabelle 10 ergibt sich dann ein phi-Koeffizient von 0,79.

1) Berechnungserleichterungen ergeben sich aufgrund der Arbeiten von GUILFORD, J. P.: The Phi Coefficient and Chi Square as Indices of Item Validity, Psychometrika 6, 1941, S. 11 - 19

Vgl. JURGENSEN, C. E.: Table for Determining Phi Coefficients, Psychometrika 12, 1947, S. 19 - 29

Nachdem die ϕ -Koeffizienten für alle Items berechnet sind, wird ein Koordinatenkreuz errichtet: Auf der Abszisse werden die Skalenwerte der Items und auf der Ordinate die Werte der entsprechenden ϕ -Koeffizienten abgetragen. EDWARDS und KILPATRICK teilten in ihrer Untersuchung die ursprünglichen neun Kategorien auf der Abszisse noch einmal in je zwei Halbkategorien auf; nur sieben dieser Halbkategorien enthielten jedoch aufgrund der errechneten Skalenwerte auch Items. Aus jeder der besetzten Halbkategorien entnahmen EDWARDS und KILPATRICK dann die vier Items mit den höchsten ϕ -Koeffizienten, der höchsten Diskriminationsfähigkeit also. Auf diese Weise wurden 28 Items ausgewählt, nach ihren Skalenwerten geordnet und zu zwei Einstellungsskalen, A und B, zusammengesetzt, wobei die Items den beiden Skalen abwechselnd zugewiesen wurden. Es entstanden also zwei Skalen, deren je 14 Items folgendermaßen charakterisierbar sind: mittlere Skalenwerte 3,85 bzw. 3,91, mittlere Q-Werte 0,90 bzw. 0,92, Mediane der ϕ -Koeffizienten 0,65 bzw. 0,66, wobei kein ϕ -Koeffizient niedriger als 0,58 ist.

Die beiden so konstruierten Skalen wurden dann erneut einer Gruppe von Versuchspersonen vorgelegt; sie äußerten ihre Zustimmung oder Ablehnung zu den einzelnen Items aufgrund der ursprünglichen sechs Antwortkategorien. Die Antworten wurden nach dem bereits beschriebenen Verfahren dichotomisiert und mit 1 bzw. 0 bewertet. Danach wurde jede der beiden Itemreihen einer Skalogramm-Analyse unterzogen. Das Ergebnis war ein Reproduktionskoeffizient von 0,875 für die Skala A und von 0,872 für die Skala B. Der mittlere Wert der modalen Antwortkategorien betrug für beide Skalen 0,57 und schwankte bei Skala A zwischen 0,51 und 0,82 und bei Skala B zwischen 0,52 und 0,67.

So können mit Hilfe der Diskriminationstechnik Skalen konstruiert werden, die folgende Kriterien aufweisen:

- a) Die Skalenwerte der Items sind über das psychologische Kontinuum, wie es bei der Technik gleicherscheinender Abstände zugrunde gelegt wird, ungefähr gleichmäßig verteilt; dabei ist es allerdings möglich, daß Items, deren Skalenwerte im mittleren Bereich des Kontinuums liegen und die nicht zwischen hohen und niedrigen Gruppen diskriminieren, durch die Techniken der Itemanalyse ausgeschieden werden.¹⁾
- b) Die Versuchspersonen weisen eine hohe Übereinstimmung bei der Einstufung der Items nach dem Verfahren gleicherscheinender Abstände auf.
- c) Die Items diskriminieren stark zwischen hoher und niedriger Gruppe; als Maß für die Diskriminationsfähigkeit wird der ϕ -Koeffizient berechnet.
- d) Die Reproduktionskoeffizienten sind im Vergleich zur minimalen marginalen Reproduzierbarkeit hoch; die Skalen genügen damit den Anforderungen des von GUTTMAN entwickelten Modells und ermöglichen die Bildung einer Rangfolge für Versuchspersonen und Items (Ordinalskalenniveau).

EDWARDS und KILPATRICK, die ihr Verfahren als eine "Synthese" der von THURSTONE, LIKERT und GUTTMAN erarbeiteten Techniken zur Prüfung von Items ansehen, behaupten, daß ihre Vorgehensweise Vorteile habe, die jede dieser Techniken für sich allein nicht besitze:²⁾ Die Technik gleicherscheinender Abstände THURSTONEs bietet keine Möglichkeit, zwischen Items mit hoher und niedriger Diskriminationsfähigkeit zu unterscheiden. Zur Aufnahme in eine Skala aber können Items,

1) Vgl. hierzu auch weiter unten am Ende dieses Abschnittes.

2) Vgl. EDWARDS, A. L. und F. P. KILPATRICK: A Technique for the Construction of Attitude Scales, a.a.O., S. 382 f.

selbst wenn sie vergleichbare Skalen- und Q-Werte haben, aufgrund ihrer verschiedenen Diskriminationsfähigkeit sehr unterschiedlich geeignet sein. Durch die Verwendung eines Maßes für Diskriminationsfähigkeit kann in dieser Frage eine Entscheidung anhand definitiver Kriterien getroffen werden. Der Vorteil der Diskriminationstechnik gegenüber dem Verfahren GUTTMANs besteht darin, daß eine "objektive Grundlage" für die Auswahl einer Reihe von Items, die einer Skalogramm-Analyse unterzogen werden soll, geschaffen wird. Zwar kann auch bei Anwendung dieses Verfahrens die Skalierung der Items fehlschlagen, EDWARDS und KILPATRICK nehmen jedoch an, daß die Skalierbarkeit der Items durch ihre Technik eher gesichert wird, als wenn die Auswahl der Items für die Skalogramm-Analyse lediglich der Intuition des jeweiligen Forschers überlassen bleibt. Sie fügen hinzu: "Furthermore, the set of items selected by the scale discrimination technique provides a wider range of content than do the intuitive Guttman items."¹⁾

Trotz der Vorteile der Diskriminationstechnik fragt es sich aber, ob der große Arbeitsaufwand, der bei dieser Technik anfällt, in einem adäquaten Verhältnis zum Erfolg steht: "Der offensichtliche Einwand gegen diese Technik ist die Umständlichkeit, mit der ein höherer Grad an Sicherheit erkaufte werden muß."²⁾ So müssen die verwendeten Items jeweils drei Gruppen von Versuchspersonen vorgelegt werden, bevor die abschließende Skalogramm-Analyse erfolgen kann. Wie die einfachen Beispiele der Items aus BOGARDUS "social distance scale"³⁾ zeigen, wird ein

1) EDWARDS, A. L. und F. P. KILPATRICK: A Technique for the Construction of Attitude Scales, a.a.O., S. 383

2) SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, a.a.O., S. 359

3) Vgl. BOGARDUS, E. S.: Measuring Social Distance, a.a.O.

solcher Arbeitsaufwand oftmals von vornherein unangebracht erscheinen. Das gilt um so mehr, als die Probleme der Validität und der Reliabilität der Messung grundsätzlich in genau demselben Maße bestehen bleiben wie bei der Skalogramm-Analyse. Zudem können auf jeder Stufe der Skalenkonstruktion zufällige Fehler das Ergebnis beeinflussen, was die Reliabilität nur beeinträchtigen kann. Aufgrund der Ausführungen über das Bedeutungskollektiv¹⁾ ergibt sich, daß die Diskriminationstechnik möglicherweise dann am erfolgreichsten eingesetzt wird, wenn die Spezifität des Bedeutungskollektivs und damit auch die Chance der Skalierbarkeit gering ist.

Zu diesen mehr pragmatischen Einwänden über die Zweckmäßigkeit der Diskriminationstechnik kommt ein weiterer Einwand von theoretischer Bedeutung: Die "Judgment"-Techniken gehen von nichtmonotonen Items aus, die Technik der summierten Einschätzungen und die Skalogramm-Analyse setzen dagegen monotone Items voraus. GREEN bemerkt in einem ähnlichen Zusammenhang: "Theoretically, a nonmonotone type of item could not be used on a Likert-type scale, and a Likert-type item could not be used on a Thurstone Scale."²⁾ Es ist aus diesem Grunde offensichtlich, daß bei der Diskriminationstechnik bestimmte Annahmen des einen oder des anderen Modells verletzt werden müssen. Dieser Umstand kann jedoch unter den in der Praxis herrschenden Bedingungen vernachlässigt werden.³⁾ Außerdem gibt es eine Überlegung, die für die Vereinbarkeit der beiden Arten von Modellen spricht: Die Folge der Anwendung von Techniken der Itemanalyse auf nichtmonotone Items wird, wie bereits ausgeführt, in der Regel darin bestehen, daß solche Items,

1) Vgl. den vorangehenden Abschnitt 433 dieser Arbeit.

2) GREEN, B. F.: Attitude Measurement, a.a.O., S. 365

3) Vgl. SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, a.a.O., S. 359

deren Skalenwerte in der Mitte des psychologischen Kontinuums liegen, ausgeschieden werden, weil sie nicht zwischen den Gruppen mit hohen und niedrigen Skalenwerten diskriminieren. Auf diese Weise werden nach EDWARDS und KILPATRICK Reliabilität und Reproduzierbarkeit der Reihe von Items, welche die endgültige Skala bilden, erhöht.¹⁾ Die Items aber, deren Skalenwerte bei der Technik gleicherscheinen - der Abstände an den Enden des psychologischen Kontinuums liegen, nehmen meist eine eher monotone Verteilungsform an, wie sie bei den Techniken der summierten Einschätzungen und der Skalogramm-Analyse unterstellt wird. Wenn demnach bei der Diskriminationstechnik Items im mittleren Bereich des Kontinuums eliminiert werden und die Items an den Endpunkten eine monotone Verteilungsform aufweisen, so kann es zu einer Vereinbarkeit der beiden Modelle kommen.²⁾

435 Die Technik der transferierten Einschätzungen

Während es sich bei der Technik der Skalogramm-Analyse um Reaktions-orientiertes Messen auf der Grundlage kategorialer Urteile handelt, ist die Technik der transferierten Einschätzungen³⁾ zwar ebenfalls ein Verfahren Reaktions-orientierten Messens, es beruht aber auf vergleichenden Urteilen.

1) Vgl. EDWARDS, A. L. und F. P. KILPATRICK: A Technique for the Construction of Attitude Scales, a.a.O., S. 383

2) Vgl. GREEN, B. F.: Attitude Measurement, a.a.O., S. 365

3) Diese Bezeichnung der Technik stammt von SCHEUCH; der englische, von SIXTL auch im Deutschen verwendete Terminus für dieses Verfahren lautet "Unfolding". Vgl. SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, a.a.O., S. 366 - 368
Vgl. SIXTL, F.: Meßmethoden in der Psychologie, S. 391 - 401

Diese Technik, die ursprünglich von COOMBS entwickelt wurde¹⁾, läßt sich "als eine Anwendung von Elementen der Skalogramm-Analyse auf Rangordnungen kennzeichnen"²⁾. Die Ausgangsdaten für die Skalierung bestehen aus einer Rangordnung der Items, die alle Versuchspersonen aufgrund ihrer eigenen Einstellungen vornehmen. Jede Versuchsperson gibt also an, welches Item ihre Einstellung am besten ausdrückt, welches am zweitbesten usw., bis zu dem Item, das der Einstellung der betreffenden Versuchsperson am wenigsten entspricht. Auf diese Weise erhält man vergleichende Urteile der Form: Versuchsperson X zieht Item i dem Item j und dieses wiederum dem Item k vor. Zum allgemeinen Ansatz Reaktions-orientierten Messens gehört die Technik der transferierten Einschätzungen deshalb, weil angenommen wird, daß die vergleichenden Urteile einer Versuchsperson sowohl durch deren eigene Einstellung als auch durch die in den Items ausgedrückte Einstellung bestimmt werden.³⁾ Bei diesem Verfahren handelt es sich ebenso wie bei der Skalogramm-Analyse um ein deterministisches Modell, "in that in the model itself no provision is made for any variation in responses above and beyond that which can be accounted for by the relation of the subject and the stimuli to the underlying attribute"⁴⁾.

1) Die erste Veröffentlichung, in der COOMBS seine Ideen zur Skalierung vortrug, war COOMBS, C. H.: Psychological Scaling without a Unit of Measurement, a.a.O. Der Aufsatz wurde in Auszügen abgedruckt in M. FISHBEIN (Hrsg.): Readings in Attitude Theory and Measurement, a.a.O., S. 117 - 124.

Vgl. auch COOMBS, C. H.: A Theory of Psychological Scaling, Engng. Res. Inst. Bull. Nr. 34, Ann Arbor 1952
Vgl. ders.: Theory and Methods of Social Measurement, a.a.O.

2) SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, a.a.O., S. 336

3) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 404

4) Ebenda

Bei der Technik der transferierten Einschätzungen geht es - ganz allgemein ausgedrückt - darum, aus den Präferenzordnungen der Versuchspersonen auf ein zugrunde liegendes Einstellungskontinuum zu schließen, auf dem sowohl die Items als auch die Versuchspersonen plaziert werden können. Es soll also "für manifeste Reaktionen ein latentes Kontinuum gesucht" werden¹⁾. COOMBS selbst spricht von "genotypischen" und "phänotypischen" Ebenen der Beschreibung²⁾. Wie die Skalogramm-Analyse, so enthält auch dieses Verfahren kaum Anleitungen für die Datensammlung selbst, wie z.B. für die Formulierung der Items; es handelt sich vielmehr um eine Technik zur Analyse bereits vorliegender Daten. Ebenso gibt es bisher kaum praktikable Vorgehensweisen für die Errechnung von Skalenwerten; eine Ausnahme bildet lediglich das UPRO-Verfahren von BANTA³⁾, das auf den Vorstellungen COOMBS aufbaut. Aus diesem Grund muß die Erörterung der Technik der transferierten Einschätzungen in dieser Arbeit auf das allgemeine Modell, das dem Verfahren zugrunde liegt, sowie auf die Grundlagen der Datenanalyse, wie sie COOMBS vorschlägt, beschränkt bleiben.⁴⁾

1) SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, a.a.O., S. 366

2) Vgl. COOMBS, C. H.: Theory and Methods of Social Measurement, a.a.O., S. 488 f. und S. 500 f.

3) Vgl. BANTA, T. J.: Social Attitudes and Response Styles, Educ. Psychol. Measmt. 21, 1961, S. 543 - 557
Vgl. auch den nachfolgenden Abschnitt 436 dieser Arbeit.

4) Die Ausführungen COOMBS zur "unfolding technique" sind zum Teil nur verständlich, wenn die von ihm entwickelte Theorie des Messens und die dort verwendete spezielle Terminologie bekannt sind. Da diese Theorie hier nicht dargestellt werden soll, folgt dieser Abschnitt im wesentlichen der auch ohne die spezielle Terminologie COOMBS verständlichen Darstellung von TORGERSON, W. S.: Theory and Methods of Scaling, S. 403 ff. - sowie den Ausführungen COOMBS in dem Aufsatz: Theory and Methods of Social Measurement, a.a.O. - Zu einer sehr ausführlichen Diskussion der Technik der transferierten Einschätzungen vgl. COOMBS, C. H.: A Theory of Data, New York, London und Sidney 1964, besonders die Kapitel 5 bis 7.

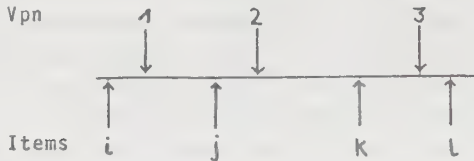
Das bei der Technik der transferierten Einschätzungen angewendete Modell erfordert, daß die Präferenzordnungen der Items, wie sie von den Versuchspersonen erstellt werden, nur auf eine einzige zugrunde liegende Variable zurückzuführen sind. Diese Variable kann in Form eines eindimensionalen Kontinuums als Abschnitt auf einer Geraden dargestellt werden. Sowohl die Versuchspersonen als auch die Items sind als Punkte auf dieser Geraden lokalisierbar. Es wird weiter angenommen, daß der Abstand jedes Items von jeder Versuchsperson direkt von dem Ausmaß abhängt, in dem eine Versuchsperson bereit ist, das betreffende Item auszuwählen bzw. diesem Item zuzustimmen. Das bedeutet: Wenn der Ort eines Items mit dem Ort einer Versuchsperson auf der Geraden zusammenfällt, so handelt es sich dabei um ein Item, dem die Versuchsperson völlig zustimmt. Je mehr sich der Ort eines Items von dem einer Versuchsperson entfernt, um so weniger wird diese Versuchsperson bereit sein, dem Item zuzustimmen. Wenn s_1 der Skalenwert von Versuchsperson 1 und s_i und s_j die Skalenwerte der Items i und j sind, dann bedeutet das Urteil der Versuchsperson 1: "Ich ziehe Item i dem Item j vor":

$$|s_i - s_1| < |s_j - s_1| \quad \text{oder} \quad d_{1i} < d_{1j}$$

wobei d der Abstand zwischen den Skalenwerten ist. Es muß betont werden, daß in dieser Betrachtungsweise die Abstände nicht gerichtet sind.¹⁾ Daraus geht hervor, daß die Präferenzordnung einer Versuchsperson der Rangordnung der Abstände der Items von der Position der betreffenden Versuchsperson auf dem

1) Es handelt sich hier um eine etwas unrealistische Betrachtungsweise, weil Aussagen über Abstände zumindest das Niveau einer geordneten metrischen Skala voraussetzen, deren Bestehen jedoch nicht ohne weiteres unterstellt werden kann.

zugrunde liegenden Kontinuum entspricht. Dabei richten sich die von den einzelnen Versuchspersonen gewählten Präferenzordnungen nach ihren eigenen Positionen auf dem Kontinuum. Das kann anhand der Figur 4 illustriert werden¹⁾:

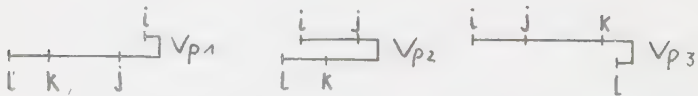


Figur 4: Konstruiertes Beispiel für die Positionen von drei Versuchspersonen (1 bis 3) und vier Items (i bis l) auf einem Kontinuum

Für die Versuchspersonen ergeben sich damit folgende Präferenzordnungen:

- Vp 1: i j k l
 Vp 2: j k i l
 Vp 3: l k j i

Das Zustandekommen dieser Präferenzordnung kann man sich graphisch so vorstellen, als ob das Kontinuum an der jeweiligen Position der Versuchspersonen um 180° gedreht ("gefaltet") würde:²⁾



Figur 5: An der Position der Versuchspersonen gefaltetes Kontinuum

- 1) Diese Figur kann zunächst nur als konstruiertes Beispiel gezeichnet werden, weil die Positionen der Items und der Versuchspersonen ja vor erfolgter Skalierung unbekannt sind.
- 2) Vgl. die Darstellung bei SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, a.a.O., S. 366

Diese an der Position der Versuchspersonen gefalteten Kontinua nennt COOMBS "qualitative I scales" oder einfach "I scales"; das entfaltete Kontinuum, das die Positionen von Items und Versuchspersonen enthält, wird "joint scale" oder einfach "J scale" genannt.

Die eigentliche Skalierungsaufgabe läßt sich nun kurz so formulieren: "The scaling task can be considered in terms of unfolding the obtained rank orders of preferences onto the single underlying continuum"¹⁾. Aus den Präferenzanordnungen der Versuchspersonen kann in einem Schritt eine Skala konstruiert werden, die sowohl die Positionen der Items als auch die der Versuchspersonen enthält. Dazu muß allerdings eine Reihe von Voraussetzungen erfüllt sein, die bei SIXTL folgendermaßen zusammengefaßt werden:²⁾

- a) Es muß eine interindividuelle Eindimensionalität vorliegen. SIXTL meint damit, daß die Relationen zwischen den verwendeten Items für alle Versuchspersonen gleich sein müssen.
- b) Die Positionen der Versuchspersonen müssen von den Positionen der Items unabhängig sein.
- c) Die Positionen der Versuchspersonen müssen ebenso wie die Positionen der Items über das gesamte Einstellungskontinuum verteilt sein.

Wenn diese Voraussetzungen gegeben sind, so enthalten die

1) TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 405
Aus diesem Grunde nennt COOMBS sein Verfahren "unfolding technique".
Vgl. COOMBS, C. H.: Psychological Scaling without a Unit of Measurement, a.a.O., abgedruckt in: M. FISHBEIN (Hrsg.): Readings in Attitude Theory and Measurement, a.a.O., S. 118
SCHEUCH leitet seine Bezeichnung der Technik als "Verfahren der transferierten Einschätzungen" von der Vorstellung her, "daß die Reaktionen jeweils auf ein subjektives Kontinuum übertragen werden".
SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, a.a.O., S. 367

2) Vgl. SIXTL, F.: Meßmethoden in der Psychologie, a.a.O., S. 392

Daten der I-Skalen eine Reihe von Informationen: "1. The discovery of a latent attribute underlying preferences. 2. The order of the stimuli on the attribute continuum. 3. Something about the relative magnitudes of the distances between pairs of stimuli. 4. The sectioning of the continuum into intervals, the placing of people in these intervals, and the ordering of these intervals on this attribute continuum. 5. Something about the relative magnitudes of these intervals"¹⁾.

Es ist möglich, daß jede Versuchsperson eine vollständige Rangordnung der Items angibt. In einem solchen Falle haben die Versuchspersonen die Aufgabe, alle n Items je nach ihren eigenen Präferenzen zu ordnen; diese Vorgehensweise setzt Transitivität und Konsistenz der Urteile der Versuchspersonen voraus²⁾. Es kann jedoch auch sein, daß die Versuchspersonen aufgrund der ihnen gestellten Aufgabe aus der Gesamtzahl der n Items nur die k Items auswählen sollen, die ihrer Einstellung am nächsten kommen. Dabei ist $k < n - 1$. Die Versuchspersonen geben dann lediglich ihre erste, zweite, dritte usw. Wahl an. Es ist offensichtlich, daß bei $k = n - 1$ die beiden Vorgehensweisen identisch werden.

Die Erläuterung der Datenanalyse bei der Technik der transferierten Einschätzungen soll hier anhand des von TORGERSON gewählten illustrativen Beispiels erfolgen.³⁾ Zunächst zum Fall der vollständigen Präferenzordnungen: Wenn eine Reihe von Ver-

1) COOMBS, C. H.: Psychological Scaling without a Unit of Measurement, a.a.O., abgedruckt in: M. FISHBEIN (Hrsg.): Readings in Attitude Theory and Measurement, a.a.O., S. 124

2) Es gibt Möglichkeiten zur Prüfung dieser Voraussetzungen, wie z.B. die Anwendung der Technik des Paarvergleichs u.ä. Hierauf soll jedoch nicht näher eingegangen werden.

3) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 406 ff.

suchspersonen ihre Präferenzen für vier Sorten von Speiseeis, die sich nur durch ihren Süßigkeitsgehalt unterscheiden, angegeben hat, so erhält man - und das sind die Ausgangsdaten der Analyse - für jede dieser Versuchspersonen eine I-Skala. Im Falle der vier Speiseeissorten gibt es $4! = 24$ mögliche Präferenzordnungen. Wenn die Daten mit dem Modell übereinstimmen, dürfen aber - wie zu begründen sein wird - nur $1/2 n (n - 1) + 1 = 7$ der insgesamt möglichen Präferenzordnungen tatsächlich auftreten. Ähnlich der Skalogramm-Analyse und dem letzten Schritt der Diskriminationstechnik muß je nach dem Auftreten abweichender Antwortschemata eine Entscheidung darüber gefällt werden, ob die Reaktionen der Versuchspersonen auf die Möglichkeit einer Ordnung von Items und Versuchspersonen auf einem eindimensionalen Kontinuum schließen lassen.

Die Mittelpunkte zwischen den Positionen zweier Items auf dem Kontinuum teilen die Versuchspersonen in jeweils zwei Kategorien ein, wie es anhand der Figur 6 deutlich wird:



Figur 6: Die Positionen der Mittelpunkte zwischen allen Itempaaren

M_{ij} bezeichnet in diesem Beispiel den Mittelpunkt der Strecke zwischen den Items i und j auf dem psychologischen Kontinuum. Alle Versuchspersonen links von M_{ij} geben die Reihenfolge ij , alle Versuchspersonen rechts von M_{ij} die Reihenfolge ji an. Die Mittelpunkte zwischen jeweils zwei Items dienen so als Grenzen zwischen geordneten Kategorien von Versuchspersonen. Werden den Versuchsperso-

nen n Items vorgelegt, so ergeben sich $1/2 n (n - 1)$ Mittelpunkte und $1/2 n (n - 1) + 1$ geordnete Kategorien von Versuchspersonen. Für das Beispiel TORGERSONs ergeben sich somit sechs Mittelpunkte und sieben Kategorien. Alle Versuchspersonen in einer bestimmten Kategorie geben dieselbe Reihenfolge für die Items an. Für die sieben Kategorien der Figur 6 erhält man dann folgende möglichen Reihenfolgen der Items:

Kategorie 1:	i j k l
" 2:	j i k l
" 3:	j k i l
" 4:	j k l i
" 5:	k j l i
" 6:	k l j i
" 7:	l k j i

Wenn also die Items die Positionen auf dem psychologischen Kontinuum haben, wie sie in Figur 6 eingetragen sind, und wenn die Präferenzordnungen der Versuchspersonen dem Modell entsprechen, so werden in empirischen Untersuchungen lediglich Rangordnungen der Form 1 bis 7 auftreten. Anhand der tatsächlich auftretenden Präferenzordnungen der Versuchspersonen wird dann versucht, die Positionen der Items auf dem psychologischen Kontinuum und die Mittelpunkte zwischen ihnen zu rekonstruieren. Dabei kann man folgendermaßen vorgehen:

Zwei Präferenzordnungen (1 und 7) bilden eine genaue Umkehrung; sie geben bereits unmittelbar die Ordnung der Items auf dem zugrunde liegenden Kontinuum wieder. Dabei kann sowohl die Präferenzordnung der Kategorie 1 als auch die der Kategorie 7 der weiteren Analyse zugrunde gelegt werden, weil das Verfahren der transferierten Einschätzungen keine Möglichkeiten zur Festlegung einer positiven oder ne-

gativen Richtung auf dem Kontinuum bietet. Ein künstlicher Nullpunkt könnte jedoch als Mittelpunkt zwischen den Items konstruiert werden, die auf dem psychologischen Kontinuum am weitesten auseinander liegen. In diesem Falle kann eine mit der Technik der transferierten Einschätzungen konstruierte Skala zur Messung von Richtung und Intensität der Einstellungen von Kommunikatoren und Rezipienten im Rahmen des erläuterten theoretischen Ansatzes verwendet werden.

Wenn auf diese Weise die Ordnung der vier Items anhand der extremen Kategorien ermittelt ist, kann die Ordnung der sechs Mittelpunkte erfolgen. Die Präferenzordnungen, die von Versuchspersonen aus angrenzenden Kategorien gegeben werden, unterscheiden sich voneinander nur durch den Tausch lediglich zweier Items in eben dieser Präferenzordnung. Die Grenze zwischen zwei angrenzenden Kategorien wird durch den Mittelpunkt zwischen diesen beiden ausgetauschten Items gebildet. Für das Beispiel der Figur 6 bedeutet das: Die Präferenzordnungen der Kategorie 1 (i j k l) und der Kategorie 2 (j i k l) unterscheiden sich nur durch den Austausch der Items i und j; infolgedessen ist der Mittelpunkt M_{ij} die Grenze zwischen den Kategorien 1 und 2. Auf dieselbe Art können die Grenzen zwischen den übrigen Kategorien bestimmt werden. Die Festlegung der Ordnung der Mittelpunkte und damit der Grenzen zwischen den Kategorien erfolgt also in der Weise, daß ausgehend von einer der beiden extremen Kategorien mit umgekehrten Präferenzordnungen diejenigen Kategorien ermittelt werden, die sich jeweils nur durch den Austausch zweier nebeneinander liegender Items unterscheiden. Ist eine solche Kategorie ermittelt, wird aus den verbleibenden wiederum diejenige ausgesucht, die sich nur durch den Austausch zweier angrenzender Items von der vorhergehenden unterscheidet. Das Verfahren wird fortgesetzt, bis man zur entgegengesetzten extremen Kategorie gelangt.

Insoweit ist also eine Ordnung der Items und der Versuchspersonen auf dem zugrunde liegenden psychologischen Kontinuum möglich; das bedeutet eine Messung auf Ordinalskalenniveau. Die Präferenzordnungen der Versuchspersonen liefern jedoch noch eine weitere Information, welche die Anhebung des Meßniveaus auf das einer geordneten metrischen Skala¹⁾ ermöglicht, da sich diese Information auf die relative Größe der Abstände zwischen den Items bezieht: Zunächst geht aus der Ordnung der Items unmittelbar hervor, daß z.B. im Falle von drei Items i, j und k der Abstand $\overline{ik}$ entweder größer als der Abstand $\overline{ij}$ oder größer als der Abstand $\overline{jk}$ sein muß. Darüber hinaus enthalten die Präferenzordnungen der Versuchspersonen noch weitere Informationen: Für vier zu skalierende Items können zwei intern konsistente Reihen von Präferenzordnungen auftauchen. Welche von beiden Reihen tatsächlich auftritt, richtet sich nach der Position der Items - im Beispiel TORGERSONs also der Speiseeissorten, die sich lediglich durch ihren Gehalt an Süßigkeiten unterscheiden - auf dem psychologischen Kontinuum. Für eine Reihe geordneter Items besteht die entscheidende Information darin, welche Lage der Mittelpunkt zwischen dem ersten und dem letzten Item im Vergleich zur Lage des Mittelpunktes zwischen den beiden mittleren Items der geordneten Reihe hat. Für die vier Items des Beispiels gilt: Geht der Mittelpunkt der extremen Items dem Mittelpunkt der beiden mittleren Items voraus, so ist der Abstand zwischen dem ersten und dem zweiten Item größer als der Abstand zwischen dem dritten und dem vierten²⁾. Das trifft für die Ordnung der Items in Form der Figur 6 zu: Weil der

1) Vgl. dazu Abschnitt 32 dieser Arbeit.

2) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 408

Mittelpunkt M_{ij} der beiden extremen Items dem Mittelpunkt M_{jk} der beiden mittleren Items vorausgeht, ist $\overline{ij} > \overline{kt}$. Ebenso mit dem Modell vereinbar wäre folgende Reihe von Präferenzordnungen:

Kategorie 1:	i j k l
" 2:	j i k l
" 3:	j k i l
" 4:	<u>k j i l</u>
" 5:	k j l i
" 6:	k l j i
" 7:	l k j i

In diesem Falle enthält die 4. Kategorie mit $k j i l$ eine vom ursprünglichen Beispiel abweichende Präferenzordnung der Items, die dort $j k l i$ lautete. Daraus geht hervor, daß nunmehr der Mittelpunkt M_{jk} dem Mittelpunkt M_{il} vorausgeht. Für die Abstände zwischen den Items gilt jetzt: $\overline{kt} > \overline{ij}$.¹⁾

Die Technik der transferierten Einschätzungen liefert also auch Informationen über die relative Größe einiger Abstände zwischen den Items. Wenn auf diese Weise nur einige Abstände zwischen nebeneinander liegenden Items geordnet werden, liegt eine Messung auf dem Niveau einer geordneten metrischen Skala vor. Ist eine Rangordnung der Abstände zwischen allen Itempaaren erreichbar, so spricht man - wie gesagt - von einer höher geordneten metrischen Skala; das bedeutet eine weitere Annäherung an das Niveau einer Intervallskala.

1) Die graphische Darstellung der Figur 6, S. 207 dieser Arbeit, verdeutlicht diese Überlegung.

Es erscheint sogar möglich, wie SIXTL es vorschlägt¹⁾, die Technik der transferierten Einschätzungen auf Intervallskalenniveau durchzuführen, aber nur dann, wenn die Skalenwerte der Items bereits mit Hilfe eines anderen Verfahrens errechnet wurden. Dabei bleibt allerdings offen, wie die Skalenwerte der Items ermittelt werden, und zwar so, daß man mit einiger Sicherheit Intervallskalenniveau annehmen kann. Wenn das jedoch unterstellt wird und die Skalenwerte der Items somit feststehen - z.B. $i = 0$, $j = 4$, $k = 6,8$ und $l = 8,3$ ²⁾ - so errechnet man zunächst die Mittelpunkte zwischen allen Itempaaren und erhält: $M_{ij} = 2$, $M_{ik} = 3,4$, $M_{il} = 4,15$, $M_{jk} = 5,4$, $M_{jl} = 6,15$ und $M_{kl} = 7,55$. Die auf diese Weise errechneten Mittelpunkte zwischen den Itempaaren bilden die Kategorien-
grenzen. Nun sucht man die Mittelpunkte der einzelnen Kategorien und gelangt so zu den Einstellungsmaßen für die zu den verschiedenen Kategorien gehörenden Versuchspersonen. Dabei lassen sich für die beiden extremen Kategorien keine Mittelpunkte berechnen³⁾; es ergibt sich dann:

Kategorie 1:	-
" 2:	2,7
" 3:	3,78
" 4:	4,78
" 5:	5,78
" 6:	6,85
" 7:	-

- 1) Vgl. SIXTL, F.: Meßmethoden in der Psychologie, a.a.O., S. 395
- 2) Das Beispiel stammt von SIXTL, F.: Meßmethoden in der Psychologie, a.a.O., S. 395
- 3) SIXTL bemerkt hierzu: "die Position von Pbn (Probanden) in einem Randintervall ist unbestimmt; wenn man eine Normalverteilung der Einstellungsmaße erhält, kann man die Abzisse der Flächenhalbierenden als Maß der Einstellung in den Randintervallen adaptieren."
SIXTL, F.: Meßmethoden in der Psychologie, a.a.O., S. 395

Da die Ordnung der Items durch ihre Skalenwerte vorgegeben ist, läßt sich die zu jeder Kategorie gehörende Präferenzordnung sehr leicht ermitteln. Man kann dann jede Versuchsperson anhand der von ihr gewählten Präferenzordnung der entsprechenden Kategorie zuordnen und erhält so den Meßwert für ihre Einstellung.

Es fragt sich jedoch, wie bereits angedeutet, ob die für eine Anwendung dieses Verfahrens notwendige Voraussetzung eines Intervallskalenniveaus gegeben ist: Dem Einsatz des von SIXTL befürworteten Verfahrens muß die Anwendung einer anderen Technik - etwa einer "Judgment"-Technik - vorausgehen. Die gegenüber diesen Techniken erhobenen Einwände bezüglich des von ihnen erreichten Skalenniveaus müssen auch an dieser Stelle geltend gemacht werden. Zudem: Weil der hier erörterten Variante der Technik der transferierten Einschätzungen die Durchführung z.B. einer der "Judgment"-Techniken vorausgehen muß, ist diese Variante eigentlich nichts anderes als eine besonders komplizierte Vorgehensweise bei der Anwendung einer Skala, die mit Hilfe einer "Judgment"-Technik konstruiert wurde. Dabei ist jedoch zweifelhaft, ob ein solcher Aufwand in einem adäquaten Verhältnis zu der dabei zu erzielenden Genauigkeit des Messens steht.

Wie oben bereits erwähnt, kann die Technik der transferierten Einschätzungen auch dann angewendet werden, wenn die Versuchspersonen *keine vollständige Präferenzordnung* der Items angeben, sondern lediglich aus der Gesamtzahl der *n* Items die *k* Items auswählen, die ihrer Einstellung am nächsten kommen. Das in diesem Fall anzuwendende Verfahren der Datenanalyse nennt COOMBS die "Parallelogrammtechnik"¹⁾. Sie besteht

1) Diese Technik kann in der vorliegenden Arbeit nur in ihren Grundzügen dargestellt werden; für eine ausführliche Erörterung vgl. vor allem: COOMBS, C.H.: A Theory of Data, a.a.O., Kap. 4, S. 65 - 79

"im Prinzip aus Regeln für die Anordnung eines Materials in einer Matrix"¹⁾ und weist große Ähnlichkeit zu dem bereits diskutierten Verfahren der Skalogramm-Analyse auf.²⁾ Wenn die Versuchspersonen die Aufgabe erhalten, aus $n = 5$ Items, die $k = 3$ Items auszuwählen, die ihrer Einstellung am ehesten entsprechen, so würde das Skalogramm - vorausgesetzt, daß keine Antwortschemata auftreten, die mit dem Modell nicht vereinbar sind - folgendermaßen aussehen:

		Items				
		i	j	k	l	m
Antwort- schemata	I	x	x	x		
	II		x	x	x	
	III			x	x	x

Tabelle 12: Skalogramm für die Auswahl von 3 aus 5 Items³⁾

Die Antworten der Versuchspersonen enthalten jedoch wesentlich mehr Informationen, wenn die ursprüngliche Aufgabe dahingehend abgewandelt wird, daß nicht nur die drei Items anzugeben sind, mit welchen die Versuchspersonen am ehesten übereinstimmen, sondern daß diese drei Items darüber hinaus geordnet werden müssen. Die Versuchspersonen sollen dann also ihre erste, zweite, dritte Wahl usw. angeben. Entsprechend diesen Wahlen werden den Items die Rangplatzziffern 1, 2, 3 usw. zugeteilt. COOMBS nennt die erste Aufgabe "Pick k " Items und die zweite

- 1) SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, a.a.O., S. 378.
- 2) Vgl. dazu Abschnitt 433 dieser Arbeit.
- 3) Die Tabelle 12 wurde in leicht veränderter Form übernommen von TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 410, Tabelle 2; ein x bedeutet eine positive Antwort einer Versuchsperson auf ein Item.

Aufgabe "Order k" Items.¹⁾ Wie bereits gesagt, entsteht für den Sonderfall, daß $k = n - 1$ ist, eine vollständige I-Skala für jede Versuchsperson.

Die Antworten der Versuchspersonen können nun in einer Matrix angeordnet werden, nur daß die Eintragungen in dieser Matrix nicht aus den Zeichen der Tabelle 12, sondern aus den Rangplatzziffern der Items bestehen; auf diese Weise entsteht ein Parallelogramm-Diagramm. Die Aufgabe der Datenanalyse besteht darin, durch Permutation der Zeilen und Spalten eines solchen Diagramms eine Matrix zu erstellen, die folgenden Bedingungen genügt:²⁾

- a) Die Eintragungen jeder Zeile und jeder Spalte müssen ohne Lücken aufeinanderfolgen.
- b) Die Eintragungen der ersten Zeile und der ersten Spalte müssen von links nach rechts bzw. von oben nach unten monoton ansteigen.
- c) Die Eintragungen der letzten Zeile und der letzten Spalte müssen von links nach rechts bzw. von oben nach unten monoton abnehmen.
- d) Die Eintragungen in allen anderen Spalten müssen von oben nach unten zunächst monoton abnehmen und dann wieder ansteigen.

Eine Matrix, die diese Anforderungen erfüllt, hat die Form der Tabelle 13:

1) Vgl. z.B. COOMBS, C.H.: Theory and Methods of Social Measurement, a.a.O., S. 513 ff. et passim

2) Vgl. COOMBS, C.H.: Theory and Methods of Social Measurement, a.a.O., S. 516

		Items				
		i	j	k	l	m
Antwort- schemata	I	1	2	3		
	II	2	1	3		
	III	3	1	2		
	IV	3	2	1		
	V		2	1	3	
	VI		3	1	2	
	VII			1	2	3
	VIII			2	1	3
	IX			3	1	2
	X			3	2	1

Tabelle 13: Parallelogramm-Diagramm für die Ordnung von 3 aus 5 Items¹⁾

Wenn es möglich ist, aus den Antworten der Versuchspersonen, die gemäß dem "Order k"-Verfahren gegeben wurden, eine solche Matrix zu erstellen, so sind damit alle notwendigen und ausreichenden Bedingungen erfüllt, um anzunehmen, daß die Präferenzordnungen der Versuchspersonen durch *e i n e n* zugrunde liegenden Faktor bestimmt sind und somit die Ordnung von Items und Versuchspersonen auf einem eindimensionalen psychologischen Kontinuum möglich ist.²⁾

- 1) Die Tabelle 13 wurde leicht geändert von COOMBS, C.H.: Theory and Methods of Social Measurement, a.a.O., S. 516, Figur 10, übernommen.
- 2) Wie bei den bereits diskutierten Techniken, so soll auch hier aus den genannten Gründen keine Erörterung der Anwendung des Verfahrens im Bereich multidimensionaler Skalierung erfolgen. Vgl. dazu u.a.: BENNETT, J. F.: Determination of the Number of Independent Parameters of a Score Matrix from the Examination of Rank Orders, Psychometrika 21, 1956, S. 383 - 395; - BENNETT, J. F. und W. L. HAYS: Multidimensional Unfolding: Determining the Dimensionality of Ranked Preference Data, Psychometrika 25, 1960, S. 27 - 43; - COOMBS, C. H. und

Das "Order k"-Verfahren weist gegenüber dem "Pick k"-Verfahren eine Reihe von Vorteilen auf, die bei TORGERSON folgendermaßen zusammengefaßt werden¹⁾: Das "Order k"-Verfahren liefert - wie bereits gesagt - ein erhebliches Mehr an Information. Bei dem "Pick k"-Verfahren können in dem oben angeführten Beispiel die Versuchspersonen lediglich drei aufeinanderfolgenden Kategorien zugeordnet werden, bei dem "Order k"-Verfahren ist dagegen eine Zuordnung zu zehn aufeinanderfolgenden Kategorien möglich.²⁾ Wenn $k \geq 3$ ist, erhält man außerdem bei dem "Order k"-Verfahren auch die bereits diskutierten Informationen über die relative Größe der Abstände zwischen den Items. Die Daten enthalten um so mehr Informationen über die relative Größe dieser Abstände, je höher der k-Wert im Verhältnis zur Gesamtzahl der Items ist.

Die Behandlung abweichender Fälle ist bei allen deterministischen Skalierungsmodellen - so auch beim Verfahren der transferierten Einschätzungen - sehr problematisch. Solche Fälle müssen als "Fehler" betrachtet werden. Für die Skalogramm-Analyse sind mehr oder weniger gute Lösungen dieses Problems erarbeitet worden³⁾; das ist bei

R. C. KAO: On a Connection between Factor Analysis and Multidimensional Unfolding, in: GULLIKSEN, H. und S. MESSICK (Hrsg.): Psychological Scaling: Theory and Applications, New York und London 1960, S. 145 - 154; - HAYS, W. L. und J. F. BENNETT: Multidimensional Unfolding: Determining Configuration from Complete Rank Order Preference Data, Psychometrika 26, 1961, S. 221 - 238; - MCELWAIN, D.-W. und J. A. KEATS: Multidimensional Unfolding: Some Geometrical Solutions, Psychometrika 26, 1961, S. 325 - 332; - COOMBS, C. H.: A Theory of Data, a.a.O., Kap. 7, S. 140 - 180

- 1) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 411 f.
- 2) Vgl. Tabelle 12, S. 214, und Tabelle 13, S. 216 dieser Arbeit.
- 3) Vgl. dazu Abschnitt 433 und die Diskussion des Reproduktionskoeffizienten als Maß für die Annäherung empirischer Daten an das der Skalogramm-Analyse zugrunde liegende Modell.

dem hier erörterten Verfahren jedoch noch nicht geschehen. Zudem ist die Aufgabe, die die Versuchspersonen erfüllen müssen, recht kompliziert: Sie müssen z.B. in manchen Fällen darüber entscheiden, ob Item i in der einen Richtung von der eigenen Position genauso weit, weiter oder nicht so weit entfernt ist wie Item j in der anderen Richtung, um die gewünschte Präferenzordnung erstellen zu können. Aus diesem Grunde ist kaum zu erwarten, daß bei empirischen Untersuchungen Reihen von Präferenzordnungen gefunden werden, die dem Modell völlig entsprechen. Bevor aber geeignete Fehlermaße oder auch Möglichkeiten der Umformung des Modells in ein Wahrscheinlichkeitsmodell entwickelt werden, dürfte die Technik der transferierten Einschätzungen und das gilt auch für ihre multidimensionale Erweiterung¹⁾ - z. Zt. noch von eher theoretischem als von praktischem Interesse sein²⁾.

Wie bei allen Reaktions-orientierten Meßverfahren, so ist auch bei der Technik der transferierten Einschätzungen die Frage der Validität der auf diese Weise entwickelten Instrumente schwer zu prüfen³⁾: "nothing in either the task set for the subject or the subsequent analytical procedures specifies the nature of the underlying attribute"⁴⁾. Mit dem Ausdruck "nature of the underlying attribute" ist hier der zu operationalisierende Begriff in seiner konstitutiven Bedeutung gemeint. Wieweit also die Definition der zugrunde liegenden Variablen (underlying attribute) in ihrer konstitutiven und in ihrer operationalen Bedeutung über-

1) Eine andere Ansicht scheint SCHEUCH zu vertreten; vgl. SCHEUCH, E. K.: Skalierung, a.a.O., S. 737

2) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 413

3) In bezug auf Reliabilitätsbestimmungen dürften dagegen keine grundsätzlich neuen Probleme auftreten, die nicht schon im Zusammenhang mit den bereits dargestellten Techniken erörtert wurden.

4) TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 412

einstimmt, bleibt weitestgehend ungeklärt. Die Technik der transferierten Einschätzungen selbst ermöglicht zunächst nur eine Ordnung der Items, der Mittelpunkte zwischen den Abständen der Items auf dem psychologischen Kontinuum und der Versuchspersonen in aufeinanderfolgende Kategorien, falls die Präferenzordnungen der Versuchspersonen mit dem Modell übereinstimmen. Über die Definition der zu messenden Variablen in ihrer konstitutiven Bedeutung ist damit nichts gesagt, denn die einzelnen Schritte dieser Technik spezifizieren die Art der zugrunde liegenden Variable nicht.¹⁾ TORGERSON sieht hierin einen der Hauptunterschiede zum Indikator-orientierten Messen: Während beim Indikator-orientierten Messen "die Art der zugrunde liegenden Variablen" durch die den Versuchspersonen gestellte Aufgabe spezifiziert werde, sei das bei den Reaktions-orientierten Meßverfahren und also auch bei der Technik der transferierten Einschätzungen nicht der Fall.

Ob hierin jedoch ein grundsätzlicher Unterschied zwischen Indikator-orientiertem und Reaktions-orientiertem Messen gesehen werden kann, wie TORGERSON anzunehmen scheint, ist fraglich: Wenn z.B. beim Indikator-orientierten Messen Items in bezug auf die in ihnen ausgedrückte Einstellung zum Gegenstand einer Aussage der Massenmedien geordnet werden sollen, so ist die Aufgabe der Versuchspersonen sicherlich stärker auf die zu messende Eigenschaft ausgerichtet als im Falle Reaktions-orientierten Messens, bei dem die Aufgabe mehr oder weniger allgemein gestellt ist. Dennoch bleibt das Problem der Validierung der kon-

1) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 412

struierten Instrumente bei beiden Ansätzen dasselbe: Die Frage, wieweit nun eigentlich ein Instrument mißt, was gemessen werden soll, muß mit einem der geschilderten Verfahren geprüft werden, wobei die Prüfung der content-validity bei der Technik der transferierten Einschätzungen keine große Rolle zu spielen braucht, weil sich die Items, die den Versuchspersonen vorgelegt werden, nicht "offensichtlich" auf die zu messende Eigenschaft beziehen müssen.¹⁾ TORGERSON betont zwar ebenfalls die Notwendigkeit der Anwendung zusätzlicher Verfahren zur Prüfung dieser Frage bei der Technik der transferierten Einschätzungen²⁾, es muß jedoch hervorgehoben werden, daß solche zusätzlichen Verfahren bei den Indikator-orientierten "Judgment"-Verfahren ebenso angewendet werden müssen.

In diesem Zusammenhang ist abschließend noch auf eine Variation der Technik der transferierten Einschätzungen hinzuweisen³⁾: Falls bei dem Versuch, mit Hilfe dieser Technik eine Skala zu konstruieren, die Items im Hinblick auf die Messung einer ganz bestimmten Eigenschaft - z.B. einer Einstellung zum Gegenstand einer Aussage der Massenmedien - formuliert sind, kann die Skalenkonstruktion erheblich erleichtert werden, wenn die Versuchspersonen veranlaßt werden, nicht nur die Präferenzordnungen in der geschilderten Weise zu erstellen, sondern darüber hinaus auch die Richtung anzugeben, die jedes Item-bezogen auf die eigene Position - einnimmt. Das "unfolding", das heißt

-
- 1) Ähnliches ist, wie bereits erwähnt, bei der Technik der summierten Einschätzungen der Fall; vgl. dazu Abschnitt 432 dieser Arbeit.
 - 2) TORGERSON bemerkt dazu, daß diese Überlegungen lediglich für die J-Skalen, nicht aber für die I-Skalen zutreffen: "The attribute corresponding to any particular I scale is simply an attribute of preference for that particular subject. It is essentially an ordinal judgment scale." Vgl. TORGERSON, W.S.: Theory and Methods of Scaling, a.a.O., S. 412
 - 3) Vgl. GREEN, B. F.: Attitude Measurement, a.a.O., S. 359 Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 413

die in der Position der Versuchsperson gefaltete Skala, ist auf diese Weise unmittelbar aus den Daten ersichtlich. Außerdem liefern die so gewonnenen Daten noch eine zusätzliche Information. Die zu den einzelnen Kategorien gehörenden Versuchspersonen können unmittelbar in zwei geordnete Unterkategorien aufgeteilt werden: Die Versuchspersonen der Kategorie, welche z.B. durch die Präferenzordnung i j k l gekennzeichnet ist, können danach unterschieden werden, ob sie das Item i als zu "positiv" oder zu "negativ" im Vergleich zur eigenen Position auffassen. Bei dieser Variation der Technik kommt auch der Prüfung der content-validity eine gewisse Bedeutung zu.

436 Die UPRO-Technik

Auf den Gedanken des von COOMBS erarbeiteten Verfahrens der transferierten Einschätzungen aufbauend, wurde die Technik der "unfolded partial rank order" (UPRO) von BANTA¹⁾ entwickelt. Während bei COOMBS jedoch keine praktikable Vorgehensweise zur Errechnung von Skalenwerten für die zu skalierenden Items aufgezeigt wird, ermöglicht das Verfahren BANTAs eine solche Zuordnung von Skalenwerten.²⁾

Dem Aufsatz, in dem BANTA sein Verfahren beschreibt, liegt eine empirische Untersuchung zugrunde, bei der drei Skalierungstechniken im Hinblick auf bestimmte Gesichtspunkte miteinander verglichen werden: die Technik der summierten Einschätzungen von LIKERT, die Technik der gleicherscheidenden Abstände von THURSTONE³⁾ und das UPRO-Verfahren. Bei

1) Vgl. BANTA, T. J.: Social Attitudes and Response Styles, a.a.O.

2) Vgl. auch SHAW, M. E. und J. M. WRIGHT: Scales for the Measurement of Attitudes, a.a.O., S. 28

3) Vgl. die Abschnitte 432 und 4312 dieser Arbeit.

dem UPRO-Verfahren besteht die Aufgabe der Versuchspersonen darin, zu den ihnen vorgelegten Items anhand der folgenden fünf Kategorien Stellung zu nehmen:¹⁾

much too unfavorable for me		my feelings	much too favorable for me	
-	-	0	+	+

Durch die Zuordnung der Items zu den Kategorien dieses Systems entsteht eine partielle Ordnung der Items, deren Bezugspunkt die mittlere Kategorie 0 ("my feelings") ist: "The resulting rank order is an unfolded one due to the nature of the middle category, which is for items representing the respondent's own position on the topic in question."²⁾ Die Zuordnung der Items zu den einzelnen Kategorien erbringt somit, wenn sie in der beschriebenen Weise erfolgt "a partial rank ordering of the items in the scale, unfolded from the respondent's own position."³⁾

Im Anschluß daran wird jeder einzelnen Antwortkategorie eine ganze Zahl zugeordnet. Als Skalenwert für jedes Item wird dann ein Mittelwert aus den Antworten aller Versuchspersonen gemäß dieser Gewichtung errechnet. Unter Verwendung dieser Skalenwerte schlägt BANTA drei Möglichkeiten zur Berechnung der Punktwerte für die einzelnen Versuchspersonen vor:

- a) ein "mittlerer Zustimmungspunktwert" ("UPRO Mean Agree score"): der mittlere Skalenwert derjenigen Items, die eine Versuchsperson in die mittlere Kategorie ("my

1) Vgl. BANTA, T. J.: Social Attitudes and Response Styles, a.a.O., S. 547

2) BANTA, T. J. ebenda

3) Ebenda

- feelings") einstuft, wird als ihr Punktwert berechnet;
- b) ein "Punktwert der summierten Einschätzungen" ("UPRO Summated Rating score"): das arithmetische Mittel aller Items, zu denen eine Versuchsperson Stellung genommen hat, wird als ihr Punktwert berechnet;
- c) ein "summierter psychometrischer Punktwert" ("UPRO Summated Psychometric score"): als Punktwert für eine Versuchsperson wird der Prozentsatz der Items, welche von ihr als zu positiv im Verhältnis zur eigenen Position auf dem Einstellungskontinuum beurteilt werden oder der Prozentsatz der Items, welche die Versuchsperson in die mittlere Kategorie ("my feelings") einstufen, berechnet. Das bedeutet eine Dichotomisierung der Antwortkategorien.

Die Interkorrelationen der nach diesen Verfahren errechneten Punktwerte sind nach BANTA recht hoch, variieren jedoch je nach Einstellungsobjekt.¹⁾ Über die Reliabilität und Validität der mit Hilfe dieses Verfahrens konstruierten Skalen läßt sich nur wenig sagen, vor allem weil es sich hier um eine noch weitestgehend unerprobte Technik handelt. Lediglich der in BANTAs Untersuchung durchgeführte

1) In diesem Abschnitt soll lediglich die von BANTA entwickelte UPRO-Technik dargestellt werden. Das weitere Anliegen seiner Untersuchung ist die Prüfung der Hypothesen, die in folgenden Sätzen formuliert sind: "If it is true that a vague referent (Einstellungsobjekt, Anm. d. Verf.) introduces response set variance, then correlations between different methods of measuring the attitude should be lower for the vague referent than for the specific referent. It also follows that response set scores should vary systematically with referent ambiguity. To the extent that the present data conform to these expectations, it can be inferred that referent ambiguity is a factor in the patterning of responses to an attitude questionnaire."
BANTA, T. J.: Social Attitudes and Response Styles, a.a.O., S. 544

Auf diese Hypothesen, die im großen und ganzen bestätigt wurden, kann in der vorliegenden Arbeit nicht näher eingegangen werden.

Vergleich mit der Technik der gleicherscheinenden Abstände und der Technik der summierten Einschätzungen liefert einige Anhaltspunkte. Danach ist die Reliabilität, wenn sie nach dem Wiederholungsverfahren ermittelt wird, relativ hoch, wobei allerdings nach der Anlage der Studie BANTAs der Einwand erhoben werden kann, daß die erste Messung die nachfolgenden beeinflussen kann. Die Einschätzung der Validität - nach dem Verfahren der *criterion validity* - ergibt ebenfalls recht positive Ergebnisse, wenn als Maßstab die Korrelationen mit den nach den Techniken der gleicherscheinenden Abstände und der summierten Einschätzungen errechneten Punktwerten herangezogen werden. Demnach lassen sich mit der UPRO-Technik relativ reliable und valide Instrumente zur Messung von Einstellungen zumindest auf dem Niveau einer teilweise geordneten Skala entwickeln. Wenn eine Anhebung des Meßniveaus auf das von Ordinalskalen gelingt, was im konkreten Falle jeweils zu prüfen wäre, dann sind mit Hilfe der UPRO-Technik auch Instrumente konstruierbar, welche zur Messung der Einstellungen verwendet werden können, wie sie im Rahmen des erörterten theoretischen Ansatzes als relevant angesehen wurden.

Gegenüber anderen Skalierungsverfahren liegt der Hauptvorteil der UPRO-Technik darin, daß bereits bei einmaliger Vorlage der zu verwendenden Items sowohl die Punktwerte dieser Items als auch die Punktwerte für die Versuchspersonen errechnet werden können. Beim Verfahren der gleicherscheinenden Abstände sind dazu zwei Schritte nötig. Bei der Technik der summierten Einschätzungen sind ebenfalls zwei Schritte nötig¹⁾, zudem lassen sich hier keine Punkt-

1) BANTA sagt zwar, daß auch bei der Technik der summierten Einschätzungen nur ein Schritt notwendig sei. Vgl. BANTA, T. J.: *Social Attitudes and Response Styles*, a.a.O., S. 554. Dabei übersieht er aber, daß die Items vor der Anwendung eines entsprechenden Instrumentes zunächst einmal einer Gruppe von Versuchspersonen vorgelegt werden müssen, um eine Itemanalyse durchführen zu können, und erst nach dieser Analyse werden die Items für die endgültige Skala zusammengestellt.

werte für die einzelnen Items errechnen, so daß die UPRO-Technik demgegenüber mehr Information bietet.¹⁾

437 Die Technik der Analyse latenter Strukturen

Die Technik der Analyse latenter Strukturen, deren Entwicklung vor allem mit dem Namen LAZARSFELD²⁾ verbunden ist, soll in der vorliegenden Arbeit nur in ihren Grundzügen erläutert werden, da es sich bei diesem Verfahren eher um ein Modell als um eine ausgebaute Skalentechnik handelt.³⁾ Trotzdem erscheint eine Erörterung der Vor-

1) Vgl. BANTA, T. J.: Social Attitudes and Response Styles, a.a.O., S. 554

2) Die Arbeiten über die Analyse latenter Strukturen wurden zu Beginn vor allem im Rahmen der Forschungsabteilung der US-Armee vorangetrieben und erfuhren des weiteren besondere Unterstützung durch die RAND-Corporation. Nicht alle Arbeiten zu diesem Thema sind deshalb allgemein zugänglich; als wichtigste Veröffentlichungen LAZARSFELDS auf diesem Gebiet sind drei umfangreiche Aufsätze sowie ein jüngst erschienenes Buch anzusehen: Vgl. LAZARSFELD, P. F.: The Logical and Mathematical Foundation of Latent Structure Analysis, in: S. A. STOFFER u.a.: Measurement and Prediction, a.a.O., S. 362 - 412; - vgl. ders.: A Conceptual Introduction to Latent Structure Analysis, in: P. F. LAZARSFELD (Hrsg.): Mathematical Thinking in the Social Sciences, Glencoe, Ill. 1954, S. 347 - 387; - vgl. ders.: Latent Structure Analysis, in: S. KOCH (Hrsg.): Psychology: A Study of a Science, Bd. 3, New York, Toronto und London 1959, S. 476 - 543; - vgl. LAZARSFELD, P. F. und N. W. HENRY: Latent Structure Analysis, Boston, New York, Atlanta, Geneva, Ill., Dallas und Palo Alto 1968

3) Vgl. SCHEUCH, E. K.: Skalierung, a.a.O., S. 737
SCHEUCH bemerkt jedoch an derselben Stelle, daß diese Technik nunmehr "durchaus in einzelne durch Großrechner jetzt praktisch verwendbare Verfahren zu übersetzen ist." Er bezieht sich dabei auf BOUDON, R.: Le modèle des classes latentes, Revue Française de Sociologie 3, 1962; - SHEPARD, R. N.: Extracting Latent Structure from Behavioral Data, Proceedings of the 1964s Symposium on Digital Computing, January 30-31, 1964, at Bell Telephone Laboratories, Holmdel.
Eine Erörterung dieser Verfahren würde jedoch über den Rahmen der vorliegenden Arbeit hinausgehen.

gehensweise LAZARSELDs gerade an dieser Stelle der Arbeit als letzte zu behandelnde Technik besonders vorteilhaft, weil sie einen expliziten Bezug auf die Problematik des Verhältnisses von Theorien- und Beobachtungssprache erlaubt, das heißt also auf die Frage nach dem Verhältnis zwischen theoretischen Konstrukten und beobachtbaren Daten. Damit wird bei dieser Technik noch einmal explizit das Validitätsproblem berührt, wobei die Terminologie LAZARSELDs der in dieser Arbeit verwendeten am nächsten kommt. Ein besonderer Vorteil des Verfahrens liegt darin, daß es sehr allgemein ist: Andere Meßtechniken können als Spezialfall der Analyse latenter Strukturen angesehen werden; das wird - wie zu zeigen ist - am deutlichsten im Verhältnis zu GUTTMANs Skalogramm-Analyse. In diesem Sinne bemerkt TORGERSON, daß das bei der Analyse latenter Strukturen verwendete Modell einen ganz allgemeinen Weg darstellt, die Probleme, denen man gewöhnlich bei der Messung von Einstellungen, Fähigkeiten und anderen Persönlichkeitsmerkmalen gegenübersteht, begrifflich zu erfassen. Das Verfahren bietet nach TORGERSON die Möglichkeit einer expliziten Formulierung der bei solchen Messungen üblicherweise notwendigen Annahmen und zeigt gleichzeitig die Wege auf, wie man die Angemessenheit dieser Annahmen prüfen kann.¹⁾

Bei dem Ansatz LAZARSELDs handelt es sich in seiner allgemeinsten Form um ein probabilistisches Modell Reaktionsorientierten Messens. Dabei wird von der folgenden grundlegenden Hypothese ausgegangen: "... there exists a set of latent classes, such that the manifest relationship between any two or more items on a questionnaire can be accounted for by the existence of these latent classes and by these alone"²⁾. Das bedeutet, wie STOUFFER in seinen zusammen-

1) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 361

2) STOUFFER, S. A.: An Overview of the Contributions to Scaling and Scale Theory, in: S. A. STOUFFER u.a.: Measurement and Prediction, a.a.O., S. 19

fassenden Überlegungen betont, daß jedes Item von zwei Faktoren beeinflusst wird: von einem Faktor, der mit der zu suchenden latenten Klasse verbunden ist, und von einem weiteren Faktor, der für jedes Item spezifisch ist. Von einem solchen itemspezifischen Faktor wird angenommen, daß er sowohl gegenüber der latenten Klasse als auch gegenüber allen anderen itemspezifischen Faktoren unabhängig ist. Damit wird bereits ein Unterschied gegenüber dem Modell GUTTMANS sichtbar: Sein Modell bezieht sich lediglich auf die manifesten Beziehungen zwischen den Items. LAZARSELD dagegen "defines an attitude as an *i n f e r e n c e* as to latent classes, tested by fitting to the manifest data an appropriate latent structure model"¹⁾. LAZARSELD behandelt Einstellungen somit bewußt als hypothetische Konstrukte, als "Dispositionstermini"²⁾, und will sie als solche messen³⁾.

Ein weiterer Unterschied gegenüber dem Modell GUTTMANS besteht darin, daß es sich bei der Analyse latenter Strukturen um ein probabilistisches, bei der Skalogramm-Analyse dagegen um ein deterministisches Modell handelt. Wahrscheinlichkeitsmodelle unterscheiden sich von deterministischen dadurch, "in that they introduce explicitly notions which account for variation in responses above and beyond that

-
- 1) STOFFER, S. A.: An Overview of the Contributions to Scaling and Scale Theory, a.a.O., S. 7
 - 2) Vgl. hierzu die Erörterung in LAZARSELD, P. F.: Latent Structure Analysis, a.a.O., S. 484 - 487
 - 3) Wenn dieser Abschnitt auch auf die Diskussion der Möglichkeiten der Messung von Einstellungen beschränkt bleibt, so sei doch bemerkt, daß die Analyse latenter Strukturen durchaus auf andere Bereiche angewendet werden kann; neben der Einstellungsmessung gibt LAZARSELD z.B. an, daß mit Hilfe dieser Technik erfaßt werden könne: "the intention of a document as inferred from certain linguistic characteristics, the moral of a group as inferred from its various performances". LAZARSELD, P. F.: Latent Structure Analysis, a.a.O., S. 478

determined by the subject and item parameters"¹⁾. Während in einem deterministischen Modell die Antworten völlig durch die mit den Versuchspersonen und den Items verbundenen Parameter "vorgeschrieben" sind, wird diese völlige Bestimmtheit bei Wahrscheinlichkeitsmodellen - und so auch bei der Analyse latenter Strukturen - durch eine Wahrscheinlichkeitsaussage ersetzt: Das Modell spezifiziert lediglich, mit welcher Wahrscheinlichkeit welche Versuchspersonen welche Antworten geben. Indem somit die Fehlerhaftigkeit jedes Items in das Modell einbezogen wird, wird der "fehlerlose" Fall GÜTTMANS zum Grenzfall.²⁾ Eine solche Fassung des Modells, die - wie noch näher zu erläutern ist - "eine Formalisierung der Beziehungen zwischen latenten Strukturen und manifesten Daten"³⁾ darstellt - hat eine Reihe von Vorteilen: Einmal werden statistische Verfahren zur Prüfung der Übereinstimmung zwischen Daten und Modell anwendbar. Durch Anwendung solcher Verfahren kann die Frage geklärt werden, ob empirisch festgestellte Daten von den Werten, die aus dem Modell abgeleitet werden können, stärker abweichen, als das durch Zufallseinwirkung allein zu erwarten wäre. Außerdem kann bei dieser Vorgehensweise die allgemeine Erfahrung nutzbar gemacht werden, daß mit einer Erhöhung der Zahl der Items ceteris paribus die Meßgenauigkeit erhöht wird, während - wie bereits erläutert - für das Modell GÜTTMANS das Gegenteil zutrifft.⁴⁾ Auch wer-

1) TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 360 f.

Weil TORGERSON die diesbezüglichen Unterschiede am besten herausarbeitet, soll hier seinen Ausführungen gefolgt werden.

2) Vgl. SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, a.a.O., S. 363 f.

3) ders., a.a.O., S. 363

4) Vgl. ders., a.a.O., S. 364

Auf die Ähnlichkeit dieses Modells mit dem "single common factor model", das auch dem von LIKERT entwickelten Verfahren der summierten Einschätzungen zugrunde gelegt werden kann, wurde bereits hingewiesen; vgl. Abschnitt 432 dieser Arbeit.

Vgl. auch die Ausführungen LAZARSELDs in: LAZARSELD, P. F.: Latent Structure Analysis, a.a.O., S. 506

den dabei, wie SCHEUCH bemerkt, die konkreten Items zur Messung des intendierten Tatbestandes austauschbar. SCHEUCH schränkt jedoch ein, daß mit Hilfe dieses Verfahrens zwar die Existenz eines latenten Faktors nachweisbar, die Deutung dieses Faktors aber dem Verfahren extern sei.¹⁾ Damit kann bereits jetzt gesagt werden, daß sich in bezug auf das Validitätsproblem keine andere Situation ergibt als bei den bereits erläuterten Techniken: Auch wenn aus den empirischen Daten ein gemeinsamer latenter Faktor extrahiert werden kann, ist damit noch nichts über die konstitutive Definition dieses Faktors als hypothetisches Konstrukt gesagt. Die Klärung dieser Frage muß einem der Verfahren zur Validitätsprüfung überlassen bleiben.

TORGERSON faßt die Situation des Forschers, der das Verfahren LAZARSELDs benutzen will, folgendermaßen zusammen:²⁾ Zunächst wird davon ausgegangen, daß für die zu erfassende Variable ein eindimensionales Kontinuum existiert und daß die Versuchspersonen gemäß einer zunächst unbekannten Wahrscheinlichkeitsverteilung über dieses Kontinuum verteilt sind. Der Forscher sammelt nun solche Items, von denen er annimmt ("feels!"), daß sie Reaktionen hervorrufen, die sich auf die zu erfassende Variable beziehen. Die Items werden dann einer Gruppe von Versuchspersonen in der Absicht vorgelegt, die damit gewonnenen Informationen dafür zu benutzen, diese Versuchspersonen und eventuell auch die verwendeten Items in Beziehung auf die zugrunde liegende Variable zu beschreiben. Für die Auswahl und/oder Formulierung der Items enthält dieses Verfahren also wie die meisten anderen bereits beschriebenen Techniken keine Hinweise. Bei der Analyse latenter Struk-

1) Vgl. SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, a.a.O., S. 364

2) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 361

turen werden zumeist dichotomische Items verwendet, eine vereinfachende, aber keineswegs notwendige Einschränkung¹⁾. Das Ausgangsmaterial für die Analyse besteht somit aus den Reaktionen von Versuchspersonen auf n dichotomische Items²⁾.

Außer der Annahme eines eindimensionalen Kontinuums, auf dem die Anordnung der Versuchspersonen nach einer unbekannten Wahrscheinlichkeitsverteilung möglich ist, beruht die Analyse latenter Strukturen noch auf zwei weiteren Annahmen³⁾, die zum Teil lediglich eine Spezifizierung des bereits Gesagten bedeuten:

- a) Die Annahme einer bestimmten Reaktionskurve ("trace line", "item operating characteristic")⁴⁾: Im Falle dichotomischer Items - das sei an dieser Stelle noch einmal wiederholt - gibt die Reaktionskurve die Wahrscheinlichkeit einer positiven Reaktion auf ein Item wieder, und zwar für Versuchspersonen, die jede beliebige Position auf dem zugrunde liegenden Kontinuum einnehmen können. Die Wahrscheinlichkeit einer positiven Reaktion auf ein Item wird somit als Funktion der Position auf dem zugrunde liegenden Kontinuum ausgedrückt. In der Figur 2 dieser Arbeit sind verschiedene mögliche Reaktionskurven aufgezeichnet⁵⁾. Bei den Kur-

1) Vgl. z.B. GREEN, B. F.: Attitude Measurement, a.a.O., S. 359

2) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 362

3) Vgl. zum folgenden: TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 362 f.

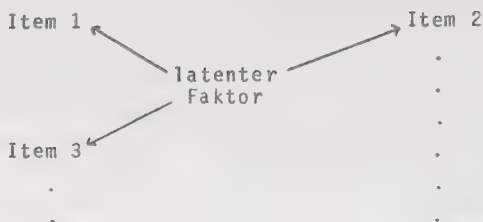
4) Vgl. hierzu die allgemeinen Ausführungen über Reaktionskurven in Abschnitt 41 dieser Arbeit. Wie gesagt, verwendet LAZARSFELD im Zusammenhang mit der Analyse latenter Strukturen den Ausdruck "trace line"; vgl. u.a. LAZARSFELD, P. F.: The Logical and Mathematical Foundation of Latent Structure Analysis, a.a.O., S. 363 ff. - LAZARSFELD, P. F.: Latent Structure Analysis, a.a.O., S. 493 f.

5) Vgl. S. 82 dieser Arbeit.

ven 1, 2 und 3 handelt es sich um Reaktionskurven, die monotonen Items entsprechen. Die Reaktionskurve 3 repräsentiert ein Item, wie es dem Modell der Skalogramm-Analyse zugrunde gelegt wird, nur daß hier die Wahrscheinlichkeit einer positiven Reaktion bis zu einem bestimmten Wert auf dem psychologischen Kontinuum etwas über null liegt und danach auch etwas unter eins bleibt. Ein Item mit dieser Reaktionskurve wäre also im Sinne des Modells bereits "fehlerhaft". Im Fall 5 liegt eine Reaktionskurve für ein nichtmonotones Item vor, wie es bei den "Judgment"-Techniken Verwendung findet. Bei der Analyse latenter Strukturen kann für jede der in der Figur 2 als Beispiele aufgeführten Reaktionskurven ein speziell angepaßtes Modell entwickelt werden. Auch hieran zeigt sich, daß sowohl die "Judgment"-Techniken als auch die Skalogramm-Analyse als Spezialfälle der Analyse latenter Strukturen aufgefaßt werden können.

- b) Die Annahme der Unabhängigkeit der Items für jeden gegebenen Punkt des zugrunde liegenden Kontinuums: Dieses Postulat wird als die wichtigste Idee der Analyse latenter Strukturen angesehen.¹⁾ Es bedeutet, daß Interkorrelationen zwischen Items ausschließlich wegen des zugrunde liegenden, latenten Faktors bestehen; das heißt, daß die genannten manifesten Interkorrelationen nur mit Hilfe des latenten Faktors und nur durch diesen zu erklären sind. Bei LAZARSELD befindet sich ein Schema²⁾, das hier in abgewandelter Form übernommen werden kann:

-
- 1) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 363
Vgl. GREEN, B. F.: Attitude Measurement, a.a.O., S. 359
- 2) Vgl. LAZARSELD, P. F.: Latent Structure Analysis, a.a.O., S. 501



Da jedes Item in irgendeiner Weise mit dem latenten Faktor verbunden ist, müssen die Items auch untereinander verbunden sein, und hierin ist der einzige Grund für ihre Interkorrelationen zu sehen: "In the latent-structure model, it is postulated that no two items have any relationship to each other above and beyond that which can be accounted for by their separate relations to the underlying variable"¹⁾. Daraus geht hervor, daß die Items bei Konstanthaltung oder Ausschaltung des latenten Faktors unabhängig voneinander variieren. Sie sind deshalb für eine Gruppe von Versuchspersonen, die eine bestimmte Position auf dem zugrunde liegenden Kontinuum einnimmt, nicht verbunden.

Die Interkorrelationen zwischen den Items können in algebraischer Formulierung des Modells der Analyse latenter Strukturen mit Hilfe der "Gleichungen der Rechnungslegung" ("accounting equations")²⁾ beschrieben werden. Um die Bedeutung dieser Gleichung zu charakterisieren, sei LAZARSELD an dieser Stelle etwas ausführlicher zitiert: "from our manifest data we actually know the frequencies

1) TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 363

2) In der vorliegenden Arbeit wird nach SCHEUCH der Ausdruck "Gleichungen der Rechnungslegung" für die Bezeichnung LAZARSELDs "accounting equations" verwendet. Vgl. SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, a.a.O., S. 363 und Fußnote 19, S. 378

in which the various response patterns occur in a given population; what we want to know are the latent parameters of the model, the coefficients which characterize the latent traces, and the distribution of the population within the latent space. We therefore need equations which link the manifest frequencies to the latent parameters. From these so-called accounting equations we then can compute all the elements in the model".¹⁾

Die Gleichungen der Rechnungslegung werden folgendermaßen aufgebaut:²⁾ Wenn x der latente Faktor ist (z.B. die Einstellung des Rezipienten zum Gegenstand einer durch ein Medium verbreiteten Aussage), dann bezeichnet $f_i(x)$ die Reaktionskurve des Items i als Funktion von x . Das bedeutet, daß $f_i(x_0)$ die Wahrscheinlichkeit einer positiven Reaktion auf das Item i für jede beliebige Position x_0 auf dem latenten Kontinuum angibt. Die Verteilung der Positionen der Versuchspersonen auf dem latenten Kontinuum wird mit $\phi(x)$ benannt. Für jedes Intervall $(x, x + dx)$ auf dem latenten Kontinuum x gilt nunmehr, daß $\phi(x) dx$ den Anteil derjenigen Versuchspersonen an der Gesamtzahl der Versuchspersonen angibt, deren latente Position in bezug auf die zu messende Einstellung in das betreffende Intervall fällt. Der Anteil von Versuchspersonen, die eine bestimmte Position auf dem latenten Kontinuum einnehmen und die auf Item i positiv reagieren, ist dann $f_i(x)\phi(x)dx$.

1) LAZARSFELD, P. F.: Latent Structure Analysis, a.a.O., S. 494

2) An dieser Stelle wird die Schreibweise benutzt, wie sie sich z.B. findet bei LAZARSFELD, P. F.: The Logical and Mathematical Foundation of Latent Structure Analysis, a.a.O., S. 368 ff. Weitere Darstellungen geben GREEN, B. F.: Attitude Measurement, a.a.O., S. 359 f., auf den sich dieser Abschnitt weitgehend stützt, sowie TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 363 f.; - SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, a.a.O., S. 364 f.

Um nun den Anteil aller auf Item i positiv reagierenden Versuchspersonen zu finden, muß über alle möglichen Werte von x integriert werden. Damit läßt sich die Gleichung 1 formulieren:

$$(1) \quad p_i = \int_{-\infty}^{+\infty} f_i(x) \phi(x) dx$$

Diese Gleichung soll erweitert werden für alle Versuchspersonen, die sowohl auf Item i als auch auf Item j positiv reagieren. Dabei gilt: Wenn die Wahrscheinlichkeiten einer positiven Reaktion auf die Items i und j mit $f_i(x)$ und $f_j(x)$ bezeichnet werden und die Items unabhängig voneinander sind, falls der latente Faktor x konstant gehalten wird, dann ist für jede beliebige Position auf dem latenten Kontinuum die Wahrscheinlichkeit einer positiven Reaktion auf beide Items i und j zugleich das Produkt $f_i(x)f_j(x)$. Daraus ergibt sich, daß der Anteil von Versuchspersonen mit einer bestimmten Position auf dem latenten Kontinuum, die auf Item i und j gleichzeitig positiv reagieren, $f_i(x)f_j(x)\phi(x)dx$ ist. Für eine Integration über alle Werte von x gilt:

$$(2) \quad p_{ij} = \int_{-\infty}^{+\infty} f_i(x)f_j(x)\phi(x)dx$$

Auf dieselbe Weise kann die Gleichung für positive Reaktionen auf beliebig viele Items erweitert werden:

$$(3) \quad p_{ijk\dots n} = \int_{-\infty}^{+\infty} f_i(x)f_j(x)f_k(x)\dots f_n(x)\phi(x)dx$$

Diese Gleichungen - eine für jedes einzelne Item, eine für jedes Itempaar, eine für jeweils drei Items usw. - beschreiben völlig die Beziehungen zwischen den Items, das heißt also die Häufigkeitsverteilung der Antwort-schemata für alle verwendeten Items.

Bei einer empirischen Untersuchung sind nun lediglich die Werte auf der linken Seite der Gleichungen beobachtbar. Die Ausdrücke auf der rechten Seite, die latenten Parameter der Items und der Verteilung der Versuchspersonen bleiben dagegen unbekannt; die Gleichungen sind demnach in der vorliegenden Form nicht lösbar. Bevor man daher überhaupt von einem Skalierungsmodell sprechen kann, müssen Einschränkungen und Modifikationen eingeführt werden, um aus den Beziehungen, wie sie in den genannten Formeln beschrieben werden, eine Reihe von Gleichungen ableiten zu können, die eine Auflösung nach den unbekannten latenten Parametern ermöglichen, wenn nur die P-Werte gegeben sind.¹⁾ Für die Einführung solcher zusätzlicher Annahmen gibt es grundsätzlich zwei Möglichkeiten:

- a) Die Annahmen können sich einmal auf die Verteilung der Positionen der Versuchspersonen auf dem latenten Kontinuum ($\phi(x)$) beziehen;
 - b) ebenso können Annahmen über die Form der Reaktionskurven der verwendeten Items ($f_i(x)$) eingeführt werden.²⁾
- Annahmen der ersten Art haben zur Entwicklung der Analyse latenter Klassen ("latent class analysis") geführt, deren einfachster Fall die Analyse latenter Dichotomien ("latent attribute analysis") ist. In diesem einfachen Fall wird von der Existenz nur zweier latenter Klassen ausgegangen: Eine wird von dem latenten Faktor beeinflusst und die andere nicht.³⁾ Durch Annahmen der zweiten Art wurde u.a. die Analyse latenter Abstände ("latent distance analysis") mit

1) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 364

2) Vgl. ebenda

3) Vgl. STOFFER, S. A.: Scaling Concepts and Scaling Theory, in: M. JAHODA, M. DEUTSCH und S. W. COOK (Hrsg.): Research Methods in Social Relations, 2. Bd., New York 1951

verschiedenen Unterfällen ermöglicht¹⁾: "In diesen verschiedenen Modellen werden die latenten Klassen usw. isoliert und dann eine Ordnung der Fälle nach ihrem Anteil an dem latenten Faktor vorgenommen."²⁾

Bei der Analyse latenter Klassen³⁾ wird davon ausgegangen, daß die Funktion $\phi(x)$ eine diskrete Verteilung bildet: Alle Versuchspersonen sind an bestimmten Punkten des latenten Kontinuums konzentriert. Die Versuchspersonen, die an ein und demselben Punkt konzentriert sind, bilden eine latente Klasse.⁴⁾ Wie findet man nun diese latenten Klassen? Gemäß dem allgemeinen Modell bei der Analyse latenter Strukturen wird angenommen, daß die Interrelationen zwischen den manifesten Items das Bestehen zweier oder mehrerer latenter Klassen widerspiegeln. Innerhalb jeder dieser latenter Klassen wird die

1) Vgl. u.a. LAZARAFELD, P. F.: The Logical and Mathematical Foundation of Latent Structure Analysis, a.a.O., S. 410 ff.; - LORD, F. M.: The Relation of Test Score to the Trait Underlying the Test, a.a.O.; - TUCKER, L. R.: Description of Paired Comparison Preference Judgments by a Multidimensional Vector Model, Educational Testing Service, Princeton, N.J., 1955

2) SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, a.a.O., S. 365.

Da die Einzelheiten der Berechnung - wie SCHEUCH an dieser Stelle bemerkt - sehr kompliziert und zum Teil nicht einmal durchgeführt worden sind, soll innerhalb der vorliegenden Arbeit lediglich die grundsätzliche Vorgehensweise bei der Analyse latenter Klassen und latenter Abstände anhand der Diskussion bei GREEN aufgezeigt werden.

Vgl. GREEN, B. F.: Attitude Measurement, a.a.O., S. 360 - 362.

Das von GREEN verwendete Beispiel wird dabei für die Zwecke dieser Arbeit leicht abgewandelt. Die von LAZARAFELD angeführten Beispiele sind in der Regel komplizierter, ohne jedoch wesentlich mehr grundlegende Informationen zu enthalten.

Vgl. z.B. LAZARSFELD, P. F.: The Interpretation and Computation of Some Latent Structures, in: S.A. STOUFFER, u.a.: Measurement and Prediction, a.a.O., S. 413 - 472

3) Vgl. auch LAZARSFELD, P. F. und N. W. HENRY: Latent Structure Analysis, a.a.O., S. 46 ff.

4) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 364

Unabhängigkeit der Items unterstellt, aber der Anteil der Versuchspersonen, die auf die einzelnen Items positiv reagieren, wird je nach Klasse verschieden sein.

Das sei an einem kurzen Beispiel erläutert, bei dem die Zahlen des von GREEN konstruierten Falles verwendet werden: Zwei Items, die sich auf das Ausmaß von politischen Sendungen im Fernsehen beziehen, sind formuliert worden:

- a) Im Fernsehprogramm sollte politischen Sendungen weit mehr Zeit eingeräumt werden, als das bisher der Fall ist.
- b) Die Sendezeit des politischen Magazins Y sollte verdoppelt werden.

Nachdem die Items a und b einem Sample von Versuchspersonen vorgelegt und die Antworten ausgezählt sind, kann sich etwa folgende Tabelle ergeben, welche den Zusammenhang zwischen den beiden Items zum Ausdruck bringt:

		Item a		insgesamt
		+	-	
Item b	+	24	23	47
	-	17	36	53
insgesamt		41	59	100

Tabelle 14: Verteilung der Antworten auf
Item a und b¹⁾

Es ist nun zu vermuten, daß das Gesamtsample aus einer Gruppe von Versuchspersonen, die an einer vermehrten politischen Information durch das Fernsehen interessiert ist, und einer Gruppe, die daran nicht interessiert ist, besteht.

1) Die Werte in der Tabelle sind Prozentzahlen, also P-Werte; dabei bedeutet + eine zustimmende und - eine ablehnende Antwort.

Da es sich hier um ein probabilistisches Modell handelt, wird nicht behauptet, daß eine klare und eindeutige Abgrenzung dieser beiden Gruppen möglich wäre, sondern lediglich, daß zwei latente Klassen mit den genannten "allgemeinen Tendenzen"¹⁾ auffindbar sind. So bemerkt LAZARSEFELD: "in latent structure analysis we do not have an empirically provided general classification against which the occurrence frequencies or probabilities of the different items can be plotted. The underlying classification is derived from the statistical behavior of the indicators themselves."²⁾ Unter der genannten Voraussetzung ist es eventuell möglich, die Werte der Tabelle 14 aufzuspalten und so tatsächlich zwei latente Klassen zu bilden: eine, in der die Versuchspersonen an vermehrter politischer Information interessiert sind, und eine, in der das nicht der Fall ist. In Tabelle 15 auf der folgenden Seite sind die Werte in dieser Weise aufgespalten.

Dieses konstruierte Beispiel läßt sich folgendermaßen interpretieren³⁾: Innerhalb jeder latenten Klasse müßten nach dem Modell die Items unabhängig sein. Das bedeutet, daß die Kenntnis der Reaktion auf ein Item keine Vorhersage der Reaktion auf ein anderes Item erlaubt. Innerhalb des Gesamtsamples muß eine positive Korrelation zwischen den Items bestehen, was auf die Existenz der latenten Klassen zurückzuführen ist. Im vorliegenden Beispiel nun sind die Items in den beiden gefundenen latenten Klassen nicht völlig unabhängig. Das liegt daran, daß außer dem Faktor "Interesse an politischer Information" noch

1) Vgl. GREEN, B. F.: Attitude Measurement, a.a.O., S. 360

2) LAZARSEFELD, P. F.: Latent Structure Analysis, a.a.O., S. 503

3) Vgl. GREEN, B. F.: Attitude Measurement, a.a.O., S. 360

an vermehrter Information
i n t e r e s s i e r t

		Item a		
		+	-	insgesamt
Item b	+	20	12	32
	-	5	3	8
insgesamt		25	15	40

+

an vermehrter Information
n i c h t i n t e r e s s i e r t

		Item a		
		+	-	insgesamt
Item b	+	4	11	15
	-	12	33	45
insgesamt		16	44	60

=

		insgesamt		
		Item b		
		+	-	insgesamt
Item b	+	24	23	47
	-	17	36	53
insgesamt		41	59	100

Tabelle 15: Aufspaltung der Werte der Tabelle 14

eine Reihe weiterer Faktoren die Reaktionen der Versuchspersonen beeinflusst. So ist es z.B. möglich, daß einer Versuchsperson zwar wenig an der Ausweitung von politischen Sendungen im Fernsehen gelegen ist; sie wird darum auf Item a ablehnend reagieren. Dieselbe Versuchsperson aber kennt den Moderator des Magazins Y persönlich und findet ihn sympatisch; aus diesem Grunde reagiert sie auf Item b positiv.

Mathematisch läßt sich das Modell der Analyse latenter Klassen in folgender Weise formulieren: Die latenten Klassen werden mit m und der Anteil der Versuchspersonen in einer bestimmten Klasse s wird mit N_s bezeichnet; die Wahrscheinlichkeit einer positiven Reaktion auf Item i in dieser Klasse s kann mit v_{is} benannt werden. Nun ist es möglich, in der Gleichung 1 den Ausdruck $\phi(x)$ durch N_s und den Ausdruck $f_i(x)$ durch v_{is} zu ersetzen. Ferner kann anstelle einer Integration über alle möglichen Werte von x eine Summation über eine endliche Zahl von Klassen erfolgen. Dann ergibt sich:

$$(4) \quad p_i = \sum_{s=1}^m v_{is} N_s$$

Bei einer Erweiterung der Formel auf ein zweites Item j gilt:

$$(5) \quad p_{ij} = \sum_{s=1}^m v_{is} v_{js} N_s$$

Für beliebig viele Items kommt man dann zu der allgemeinen Formel:

$$(6) \quad p_{ijk...n} = \sum_{s=1}^m v_{is} v_{js} v_{ks} \cdots v_{ns} N_s$$

Es sind verschiedene Verfahren entwickelt worden, um die latenten Parameter N_s und v_{is} bis v_{ns} zu ermitteln.¹⁾ Diese Verfahren sind zum Teil recht kompliziert, besonders wenn mehr als zwei oder drei latente Klassen vorhanden sind.

Die Schwächen des Modells der Analyse latenter Klassen liegen auf der Hand.²⁾ Mit der Annahme von der Konzentration der Versuchspersonen in m Klassen muß die Vorstellung eines zugrunde liegenden eindimensionalen Kontinuums aufgegeben werden. Die Summationsgleichungen 4 bis 6 sagen nichts darüber aus, daß die Klassen in irgendeiner Weise geordnet wären. Es handelt sich hier vielmehr um eine Klassifikation in Kategorien oder Typen, wobei die Parameter N_s und v_{is} eher dazu dienen, die Eigenschaften dieser Kategorien zu beschreiben, als Skalenwerte für die Versuchspersonen oder die verwendeten Items in Beziehung auf ein zugrunde liegendes Kontinuum zu errechnen. In diesem Sinne spricht GREEN davon, daß das Modell von der Vorstellung einer nominalen Einstellungsskala ausgehe, bei dem jede Versuchsperson einer der latenten Klassen zugewiesen werden könne. Hinzu kommt noch eine weitere Schwierigkeit, die nach GREEN dem Modell als logische Schwierigkeit inhärent ist: Wenn die Daten dem Mo-

1) Nicht alle Arbeiten zu diesem Problem sind veröffentlicht oder allgemein zugänglich. An dieser Stelle sei lediglich auf drei Aufsätze aus Psychometrika hingewiesen: GREEN, B. F.: A General Solution for the Latent Class Model of Latent Structure Analysis, Psychometrika 16, 1951, S. 151 - 166; - ANDERSON, T. W.: On Estimation of Parameters in Latent Structure Analysis, Psychometrika 19, 1954, S. 1 - 11; - GIBSON, W. A.: An Extension of Anderson's Solution for the Latent Structure Equations, Psychometrika 20, 1955, S. 69 - 73; - LAZARSELD, P. F. und N. W. HENRY: Latent Structure Analysis, a.a.O., S. 71 ff.

2) Vgl.: TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 365

dell adäquat sind, so kann man daraus lediglich folgern, daß eine Anzahl latenter Klassen mit bestimmten Eigenschaften besteht. Damit ist jedoch noch nicht gesagt, daß ein konkretes Sample von Versuchspersonen auch in Klassen mit diesen Eigenschaften eingeteilt werden kann.¹⁾ TORGER-SON zählt das bei der Analyse latenter Klassen verwendete Modell nicht zu den eigentlichen Skalierungstechniken. Es kann daher festgestellt werden, daß die zu diesem Modell führenden Annahmen über die Verteilung der Versuchspersonen auf dem latenten Kontinuum eine Lösung bei der Analyse latenter Strukturen ergeben, die keine Konstruktion von Instrumenten zur Messung von Einstellungen hinsichtlich Richtung und Intensität ermöglichen, es sei denn, man betrachte das Modell mit seinen diskreten Klassen als Annäherung an ein Modell, bei dem ein Kontinuum zugrunde gelegt wird.²⁾

Bei dem Modell der Analyse latenter Abstände wird von einer Annahme der zweiten Art - also über die Reaktionskurven der Items - ausgegangen: Es werden ähnliche Reaktionskurven wie bei der Skalogramm-Analyse unterstellt. Ohne auch hier auf die Einzelheiten der Berechnung eingehen zu können, kann das Modell in seinen Grundzügen in folgender Weise charakterisiert werden:³⁾ Wie bereits an anderer

1) Vgl. GREEN, B. F.: Attitude Measurement, a.a.O., S. 361

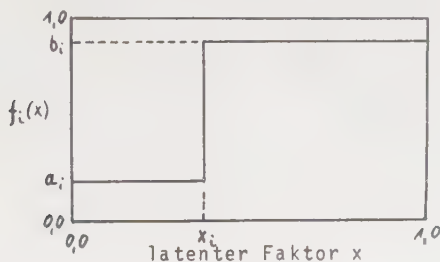
2) Vgl. ebenda

LAZARSELD erreicht in einem Beispiel eine Ordnung der Klassen durch ein "outside continuum", das mit Hilfe der Technik der summierten Einschätzungen unter Verwendung von acht Items erstellt wird. Die Ordnung der Klassen ist jedoch in diesem Falle dem Verfahren extern. Es wird dann eigentlich nur geprüft, ob der mit der Technik der summierten Einschätzungen gebildete "general attitude score" dazu benutzt werden kann, bei der Verwendung weiterer zwei Items Klassen zu bilden und die Daten im Sinne des Modells zu analysieren.
Vgl. LAZARSELD, P. F.: Latent Structure Analysis, a.a.O., S. 495 ff.

3) Vgl. neuerdings LAZARSELD, P. F. und N. W. HENRY: Latent Structure Analysis, a.a.O., S. 123 ff.

Stelle ausgeführt, handelt es sich um eine "Wahrscheinlichkeits-Analogie" zur Skalogramm-Analyse¹⁾. TORGERSON bemerkt dazu, daß eine solche Wahrscheinlichkeits-Analogie nicht nur im Verhältnis zur Skalogramm-Analyse vorliege, sondern daß jedes Modell, welches von monotonen Reaktionskurven ausgehe und Items mit hinreichender Diskriminationsfähigkeit verlange, so betrachtet werden könne.²⁾

Zu den Annahmen, wie sie bei den allgemeinen Gleichungen der Rechnungslegung gemacht wurden, kommt bei der Analyse latenter Abstände also noch ein weiteres Postulat über die Form der Reaktionskurven der verwendeten Items. Es wird angenommen, daß die Wahrscheinlichkeit einer positiven Reaktion für jedes Item bis zu einem bestimmten Punkt x_i (Wendepunkt) einen niedrigen konstanten Wert (a_i) annimmt. An diesem Punkt x_i erhöht sich die Wahrscheinlichkeit einer positiven Reaktion dann auf einen anderen Wert (b_i) und bleibt von da an wiederum konstant. Die damit unterstellte Form der Reaktionskurve wird in Figur 7 gezeigt.



Figur 7: Reaktionskurve für das Modell der Analyse latenter Abstände³⁾

- 1) Vgl. SCHEUCH, E. K.: Skalierungsverfahren in der Sozialforschung, a.a.O., S. 365
Über das Verhältnis der Analyse latenter Abstände zur Skalogramm-Analyse vgl. insbesondere LAZARSELD, P. F.: The Interpretation and Computation of Some Latent Structures, a.a.O., S. 455 - 460
- 2) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 374
- 3) Die Darstellung folgt an dieser Stelle wiederum

Für die Form der Darstellung in Figur 7 wird der Einfachheit halber unterstellt, daß die zu messende Variable x sich innerhalb des Intervalls $0 \leq x \leq 1$ bewegt; für die Verteilung der Positionen der Versuchspersonen wird $\phi(x) = 1$ angenommen. Dann kann die zu Item i gehörige Reaktionskurve mittels der folgenden Gleichung beschrieben werden:¹⁾

$$(7) \quad \begin{aligned} \text{für } x \leq x_i \text{ gilt: } f_i(x) &= a_i \\ \text{für } x > x_i \text{ gilt: } f_i(x) &= b_i \end{aligned}$$

Jedes Item hat somit drei Parameter: a_i , b_i und x_i .

Anhand dieser Überlegungen werden die Ähnlichkeiten zur Skalogramm-Analyse GUTTMANS besonders deutlich: Bei einem für die Skalogramm-Analyse perfekten Item ist $a_i = 0$ und $b_i = 1$. Der Vorteil des Modells der Analyse latenter Abstände besteht darin, daß die nicht-skalierbaren Typen GUTTMANS in diesem Falle nicht als Fehler behandelt zu werden brauchen: Jedes der Antwortschemata ist mit einem bestimmten Wahrscheinlichkeitsgrad verbunden. Zudem werden gewisse Kriterien zur Beurteilung einer mittels Skalogramm-Analyse erstellten Skala - einschließlich des Reproduktionskoeffizienten - überflüssig: "For the latent distance scale, we merely wish to know whether the model fits the data within the limits of statistical fluctuation"²⁾.

GREEN, B. F.: Attitude Measurement, a.a.O., S. 361 f.
Die Figur 7 wurde in vereinfachter Form von GREEN,
Figur 6, übernommen.

Vgl. auch die Reaktionskurve 3 der Figur 2 auf
S. 82 dieser Arbeit.

1) Diese Schreibweise entspricht der von GREEN verwendeten. Vgl. GREEN, B. F.: Attitude Measurement, a.a.O., S. 361. An anderen Stellen der Literatur werden davon abweichende Darstellungsweisen benutzt; vgl. z.B.: LAZARSELD, P. F.: The Logical and Mathematical Foundation of Latent Structure Analysis, a.a.O., S. 410 ff.; - TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 374 ff.

2) GREEN, B. F.: Attitude Measurement, a.a.O., S. 362

Zur Abschätzung der Itemparameter, die dann eine Ordnung der Items und der Versuchspersonen auf dem zugrunde liegenden Kontinuum ermöglichen, sind verschiedene Verfahren entwickelt worden, die darzustellen im Rahmen dieser Arbeit nicht notwendig ist.¹⁾ Zum einen kam es hier lediglich darauf an, die Analyse latenter Abstände in ihren Grundzügen kennenzulernen, um sich ein Bild über ihre Verwendungsmöglichkeiten machen zu können. Zum anderen ist die Lage der Massenkommunikationsforschung so, daß in der nächsten Zukunft der Einsatz einer der Techniken der Analyse latenter Abstände wegen des damit verbundenen Aufwandes kaum möglich sein wird. Das gilt auch für die Prüfung der Hypothesen des hier geschilderten theoretischen Ansatzes. An dieser Stelle sei jedoch noch betont, daß eine Ordnung der Items mit Hilfe der zu ermittelnden Wendepunkte und einiger weiterer Informationen möglich ist.²⁾ Für die Ordnung der Versuchspersonen schlägt TORGERSON in diesem Zusammenhang eine einfache Lösung vor³⁾, da andere entwickelte Verfahren sehr aufwendig sind: Danach ist eine Ordnung der Versuchspersonen auf der Grundlage der positiv beantworteten Items vorzunehmen. Wenn alle Items monoton und gleichgerichtet sind - das heißt, es muß gelten: $b_i > 0$ für alle Werte von x - dann kann ein auf diese Weise errechneter Gesamtpunktwert als Rangordnung der Versuchspersonen auf dem zugrunde liegenden Kontinuum benutzt werden.

-
- 1) Verfahren zur Abschätzung der Itemparameter sind z.B. in folgenden Arbeiten veröffentlicht worden:
LAZARSFELD, P. F.: The Interpretation and Computation of Some Latent Structures, a.a.O., besonders S. 441 - 446; - HAYS, D. G. und E. F. BORGATTA: An Empirical Comparison of Restricted and General Latent Distance Analysis, Psychometrika 19, 1954, S. 271 - 279
 - 2) Vgl. TORGERSON, W. S.: Theory and Methods of Scaling, a.a.O., S. 385
 - 3) Vgl. ders., a.a.O., S. 385

LAZARSELD faßt die Arbeitsgänge, die bei jeder Analyse latenter Strukturen durchgeführt werden müssen, in neun Schritten zusammen.¹⁾ Auch wenn die einzelnen Schritte dieser Vorgehensweise nicht eingehend diskutiert werden konnten²⁾, so ist doch die folgende Zusammenfassung gerade zum Abschluß der Diskussion der Skalierungstechniken besonders wertvoll: Bei jeder Entwicklung einer Skala müssen die von LAZARSELD genannten neun Arbeitsgänge in mehr oder weniger stark geänderter Form ausgeführt werden, wenn man Wert darauf legt, das der jeweiligen Technik zugrunde liegende Modell explizit zu formulieren. Ein Vergleich mit den bereits erläuterten Skalierungstechniken zeigt, daß dies in der Regel nicht der Fall ist. Das verdeutlicht einerseits die diesbezüglichen Mängel verschiedener Verfahren der Skalenkonstruktion und andererseits die Allgemeinheit des von LAZARSELD gewählten Ansatzes. Die ersten fünf Schritte dienen der Konstruktion eines Modells, von dem man annehmen kann, daß es den Daten im jeweils vorliegenden Fall entspricht. Erst vom sechsten Schritt an beginnt die Gegenüberstellung von Daten und Modell:

1. Zunächst ist, wie gesagt, für den jeweils vorliegenden Fall ein adäquates Modell zu formulieren. In einem solchen Modell müssen sich die Interrelationen zwischen manifesten und latenten Variablen in Form von Interrelationen zwischen den Indikatoren und dem

1) Vgl. LAZARSELD, P. F.: Latent Structure Analysis, a.a.O., S. 506 - 508.

In dieser Zusammenfassung wird die Terminologie LAZARSELDs, wenn nötig, der in dieser Arbeit verwendeten Terminologie angepaßt.

2) Wie bei allen anderen Techniken, so mußte auch hier auf Beispiele weitgehend verzichtet werden. LAZARSELD führt in seinem Aufsatz die Analyse an einem Beispiel über alle neun Stufen aus.

Vgl. LAZARSELD, P. F.: Latent Structure Analysis, a.a.O., S. 508 - 527

Er geht dabei jedoch von einer anderen Form der Reaktionskurve, als sie z.B. im Zusammenhang mit der Analyse latenter Abstände unterstellt wurde, aus.

zu operationalisierenden Begriff in seiner konstitutiven Bedeutung widerspiegeln. Die dabei zugrunde gelegten Annahmen sind explizit in mathematischer Form zu formulieren ("choice and specification of the model").

2. Sodann können die Gleichungen der Rechnungslegung, welche die Beziehungen zwischen den manifesten und den latenten Parametern wiedergeben sollen, in die von dem gewählten Modell geforderte Form gebracht werden ("accounting equations, specialized for the model").
3. Als nächstes stellt sich das Problem der Bedingungen oder Einschränkungen, die aufgrund der Annahmen des Modells für die Interrelationen zwischen den Daten gelten müssen. Diese "conditions of reducibility" sind Aussagen über die Beziehungen, die zwischen den manifesten Parametern vorliegen müssen. Durch Formulierung dieser Bedingungen kann man mit Hilfe einfacher Operationen und ohne die Gleichungen der Rechnungslegung zu lösen, etwas zur Klärung des Problems beitragen, ob das Modell für die vorliegenden Daten angemessen ist. Des weiteren kann abgeschätzt werden, wieweit überhaupt die Daten mit den Anforderungen des Modells übereinstimmen können. Außerdem tragen die "conditions of reducibility" nach LAZARSELD zum Verständnis des verwendeten Modells sowie zur Beantwortung der Frage bei, ob die Gleichungen der Rechnungslegung überhaupt lösbar sind.
4. Für den Fall, daß die manifesten Parameter vorliegen, ist dann die Frage zu klären, ob das Modell ausreicht, um die latenten Parameter zu identifizieren ("identifiability").

5. Daran schließt sich die Frage an: Wenn die Gleichungen lösbar sind, wie sollen sie im konkreten Fall gelöst werden ("identification")?
6. Wurden die ersten fünf Schritte gewissermaßen für den Idealfall einer unendlichen Samplegröße getan, so müssen nunmehr die im konkreten Fall vorliegenden Daten in die Beobachtung einbezogen werden ("computation: the fitting procedure"): LAZARSELD bemerkt dazu: "A 'fitting procedure' in latent structure analysis is usually a shuttling back and forth between data and latent parameters."¹⁾ Und weiter: "In empirical work it is from data beclouded by sampling variability that we have to find the latent parameters."²⁾ Mit der Vorgehensweise bei diesem Anpassungsprozeß werden zwei Ziele verfolgt: a) Eine Reihe latenter Parameter und b) eine Reihe "angepaßter manifester Parameter", die einerseits völlig mit den Anforderungen des Modells und andererseits so weit wie möglich mit den vorliegenden Daten übereinstimmen, sollen gefunden werden.
7. Nun ist zu prüfen, wieweit dieser Anpassungsprozeß gelungen ist ("evaluation of the fit"). Dabei sind zwei Fragen zu beantworten: a) Sind die Unterschiede zwischen den tatsächlichen und den angepaßten Parametern gering genug? b) Sind die Unterschiede zufällig verteilt oder bilden sie ein bestimmtes "Muster", so daß das Modell für den vorliegenden Fall zu modifizieren ist? Wenn sich eine Modifizierung als nötig erweist, muß mit der gesamten Analyse wieder von vorne angefangen werden, wobei die Ergebnisse der gewonnenen Erfahrungen zu berücksichtigen sind.

1) LAZARSELD, P. F.: Latent Structure Analysis, a.a.O., S. 507

2) ders., a.a.O., S. 518

8. Sind die Ergebnisse jedoch zufriedenstellend, so kann ermittelt werden, wie Versuchspersonen mit bestimmten Antwortschemata über das latente Kontinuum verteilt sind ("recruitment pattern").
9. Aufgrund dieser Ergebnisse kann schließlich festgestellt werden, wo Versuchspersonen mit bestimmten Antwortschemata auf dem latenten Kontinuum zu lokalisieren sind ("classification and scores"). Damit verbunden ist natürlich die Frage nach dem Punktwert für jede Versuchsperson bzw. für jedes Antwortschema. Ein Zusatzproblem ist die Skalierung der Items: Gegebenenfalls kann man jedem Item einen Punktwert zuschreiben, der die Diskriminationsfähigkeit dieses Items für Versuchspersonen an verschiedenen Punkten des latenten Kontinuums ausdrückt.

5 Einige Bemerkungen zum gegenwärtigen Stand der Konstruktion von Einstellungsskalen

Zum Abschluß sollen thesenartig noch einige Bemerkungen zum gegenwärtigen Stand der Meßverfahren für Einstellungen in den Sozialwissenschaften angefügt werden, wie sie anhand der bereits mehrfach zitierten Sammlung von Einstellungsskalen bei SHAW und WRIGHT formuliert werden können:¹⁾

- a) Seit den Arbeiten von THURSTONE und CHAVE, LIKERT und GUTTMAN sind auf dem Gebiet der Skalenkonstruktion trotz der großen Vielfalt von Veröffentlichungen nur wenige wirklich bedeutende Fortschritte in meßtechnischer Hinsicht erzielt worden.²⁾ Die neueren Ansätze von COOMBS und LAZARSFELD sind noch nicht voll entwickelt und spielen im Forschungsbetrieb vor allem in Deutschland bisher kaum eine Rolle.
- b) Nur wenige Fortschritte auch sind erkennbar in Richtung auf die Messung der strukturellen Charakteristika von Einstellungen und Einstellungssystemen, trotz der theoretischen Bemühungen in dieser Hinsicht.³⁾ Obwohl Einstellungen auf verschiedenen Dimensionen variieren kon-

- 1) Vgl. SHAW, M. E. und J. M. WRIGHT: Scales for the Measurement of Attitudes, a.a.O., S. 559 - 570. Die Sammlung von SHAW und WRIGHT dürfte den besten Überblick über die z.Zt. verwendeten Einstellungsskalen vermitteln. Eine weitere Zusammenstellung verschiedenster Skalen findet sich bei MILLER, D. C.: Handbook of Research Design and Social Measurement, New York 1964.
- 2) Das zeigt auch der jüngst erschienene Aufsatz von SCOTT, W. A.: Attitude Measurement, in: G. LINDZEY und L. ARONSON (Hrsg.): The Handbook of Social Psychology, 2. Aufl., Bd. 2, Reading, Mass., Menlo Park, Calif., London und Don Mills, Ontario, 1968, S. 204 - 273.
- 3) Vgl. ROTH, E.: Einstellung als Determination individuellen Verhaltens, a.a.O., S. 62 - 94.

nen, messen die bisher entwickelten Skalen im wesentlichen nur Richtung und Intensität von Einstellungen. Versuche, solche Dimensionen wie z.B. Zentralität, Komplexität und Konsistenz von Einstellungen zu messen, fehlen bisher weitgehend.

- c) Die meisten in empirischen Untersuchungen benutzten Instrumente zur Einstellungsmessung sind nicht anhand der besten zur Verfügung stehenden Skalierungsmodelle entwickelt worden; das gilt auch für die Massenkommunikationsforschung. Wie SHAW und WRIGHT feststellen, begnügt man sich häufig mit einigen nicht standardisierten Fragen und nimmt an, damit die betreffende Einstellung reliabel und valide gemessen zu haben.
- d) Für die in der Sammlung von SHAW und WRIGHT enthaltenen Skalen ist zwar immer eine bestimmte minimale Reliabilität nachgewiesen; in der fehlenden Validität vermuten die Autoren jedoch den größten Mangel der aufgenommenen Skalen.
- e) Bei der Konstruktion von Skalen in der Sozialforschung ist man sich oftmals nicht über die dem jeweils verwendeten Modell zugrunde liegenden Annahmen im klaren. So kann z.B. Unklarheit über das mit einem bestimmten Verfahren erreichte Meßniveau bestehen. Das führt dann eventuell dazu, daß Meßoperationen, die eigentlich nur auf Intervall- oder Ratioskalenniveau durchgeführt werden können, bereits auf dem Niveau einer Ordinalskala angewendet werden.
- f) Die bisher entwickelten Einstellungsskalen decken bei weitem nicht das Universum möglicher Einstellungsobjekte ab. Diese Unvollständigkeit wirkt sich besonders störend auf die Erforschung der Entstehung und Wirkung von Aussagen der Massenkommunikation aus, weil in diesen Aussagen zu den verschiedensten Einstellungsobjek-

ten Stellung genommen wird, und Skalen, die bereits auf ihre Reliabilität und Validität geprüft wurden, keineswegs in ausreichender Zahl zur Verfügung stehen.

- g) Die in die Sammlung von SHAW und WRIGHT aufgenommenen Skalen haben einen besonders schwerwiegenden Nachteil gemeinsam:¹⁾ Sie reichen nur für solche experimentellen Untersuchungen aus, in denen Einstellungen verschiedener Gruppen von Versuchspersonen verglichen werden sollen. Man nimmt dann an, daß die Meßfehler zufällig um einen Mittelwert streuen. Eine Skala, welche die Einstellung eines einzelnen Individuums nur fehlerhaft mißt, kann so möglicherweise zu einer ausreichend reliablen und validen Messung der "mittleren" Einstellung einer Gruppe von Versuchspersonen führen. Wenn jedoch eine bestimmte Einstellung nur einer Versuchsperson gemessen werden soll, ist ein Fehlerausgleich dieser Art nicht möglich. Deswegen bezweifeln SHAW und WRIGHT die Verwendbarkeit der meisten Skalen für diesen Fall.
- h) Ein besonderes Problem, auf das im Rahmen dieser Arbeit nicht eingegangen werden konnte, besteht in der Formulierung von Items. Die wesentlichen Arbeiten auf dem Gebiet der Skalenkonstruktion zur Messung von Einstellungen sind in englischer Sprache angefertigt worden; ebenso sind die Items der bekanntesten Skalen in englischer Sprache verfaßt. Eine Übertragung ins Deutsche erweist sich als außerordentlich schwierig, eine einfache Übernahme ist z.B. wegen der unterschiedlichen Wortbedeutungen und der häufig verwendeten Idiome nicht möglich. Oft können die Items der in englischer Sprache verfaßten Skalen nicht viel mehr als Anregungen für die Formulierung von Items in deutscher Sprache liefern.

1) Vgl. SHAW, M. E. und J. M. WRIGHT: Scales for the Measurement of Attitudes, a.a.O., S. 560 und 565

- i) Um die Weiterentwicklung und Prüfung einer Theorie zur Erklärung von Massenkommunikationsprozessen, deren Basisbegriff die Einstellung ist, voranzubringen, muß in der Zukunft vor allem an der Konstruktion von Skalen gearbeitet werden, die eine grundlegende Messung der Einstellungen von Kommunikatoren und Rezipienten ermöglichen. Voraussetzung dafür ist, daß man sich mehr als bisher mit Reliabilitäts- und Validitätstests beschäftigt. In engem Zusammenhang damit sind die Arbeiten an den Techniken der Aussagenanalyse zur Messung von Einstellungen, die in den Aussagen der Massenmedien geäußert werden, ebenso fortzuführen wie die Bemühungen um die Theorienbildung.

In der vorliegenden Arbeit konnte gezeigt werden, daß verschiedene Techniken zur Verfügung stehen, welche die Messung der Einstellungen von Kommunikatoren und Rezipienten auf dem Niveau von Ordinalskalen erlauben. Damit ist die diesbezügliche meßtechnische Voraussetzung zur Prüfung von Hypothesen im Rahmen des geschilderten theoretischen Ansatzes zur Erklärung der Entstehung und Wirkung von Aussagen der Massenmedien erfüllt.

LITERATURVERZEICHNIS

Abkürzungen:

Amer. Polit. Sci. Rev.	= American Political Science Review
Amer. Psychol.	= American Psychologist
Amer. Soc. Rev.	= American Sociological Review
Am. J. Psychol.	= American Journal of Psychology
Arch. Psychol.	= Archives of Psychology
Educ. Psychol. Measmt.	= Educational and Psychological Measurement
Engng. Res. Inst. Bull.	= Engineering Research Institute Bulletin
Int. J. Opinion and Attitude Res.	= International Journal of Opinion and Attitude Research
J. Abn. Soc. Psychol.	= Journal of Abnormal and Social Psychology
J. Appl. Psychol.	= Journal of Applied Psychology
J. Appl. Sociol.	= Journal of Applied Sociology
J. Educ. Psychol.	= Journal of Educational Psychology
J. Educ. Res.	= Journal of Educational Research
J. Exp. Psychol.	= Journal of Experimental Psychology
J. Gen. Psychol.	= Journal of Genetic Psychology
J. Pers. Soc. Psychol.	= Journal of Personality and Social Psychology
J. Social Issues	= Journal of Social Issues
J. Soc. Psychol.	= Journal of Social Psychology
Psychol. Bull.	= Psychological Bulletin
Psychol. Rev.	= Psychological Review

- Adorno, T. W.
Frenkel-Brunswik, E.
Levinson, D. J.
Sanford, R. N. The Authoritarian Personality,
New York 1950
- Ake, J. N. A Rapid Machine Procedure for
Determining Scalability of Any
Number of Questions, Public Opinion
Quarterly 26, 1962
- Albert, H. Probleme der Theoriebildung, in:
H. Albert (Hrsg.): Theorie und
Realität, Tübingen 1964
- American Psychological
Association Committee Technical Recommendations for
Psychological Tests and Diagnostic
Techniques, Psychol. Bull.
Supplement 51, 1954
- Anderson, T. W. On Estimation of Parameters in Latent
Structure Analysis, Psychometrika 19, 1954
- Attneave, F. A Method of Graded Dichotomies for
the Scaling of Judgments, Psychol.
Rev. 56, 1949
- Ballin, M.
Farnsworth, P. R. A Graphic Rating Method for
Determining the Scale Values of
Statements in Measuring Social Attitudes,
J. Soc. Psychol. 13, 1941
- Banta, T. J. Social Attitudes and Response Styles,
Educ. Psychol. Measmt. 21, 1961
- Bennett, J. F. Determination of the Number of
Independent Parameters of a Score
Matrix from the Examination of
Rank Orders, Psychometrika 21, 1956
- Bennett, J. F.
Hays, W. L. Multidimensional Unfolding:
Determining the Dimensionality of Ranked
Preference Data, Psychometrika 25,
1960
- Berelson, B.
Steiner, G. A. Human Behavior, New York, Chicago
und Burlingame 1964
- Bessler, H. Zur Theoriebildung in der Massenkommunikationsforschung. Ein theoretischer Ansatz und Möglichkeiten der Lösung eines damit verbundenen Meßproblems mit Techniken der Aussagenanalyse, Dissertationsdruck, Nürnberg 1968

- Bessler, H. Aussagenanalyse. Die Messung von Einstellungen im Text der Aussagen von Massenmedien, unveröffentlichter Forschungsbericht, Nürnberg 1969
- Bessler, H.
Bledjian, F. Systematik der Massenkommunikationsforschung, München und Basel 1967
- Beyle, H. C. A Scale for the Measurement of Attitude toward Candidates for Elective Governmental Office, Amer. Polit. Sci. Rev. 26, 1932
- Bird, C. Social Psychology, New York 1940
- Blalock, H. M. Jr. Social Statistics, New York, Toronto und London 1960
- Blalock, H. M. Jr. The Measurement Problem: A Gap between the Languages of Theory and Research, in: H. M. Blalock, Jr. und A. B. Blalock (Hrsg.): Methodology in Social Research, New York, St. Louis, San Francisco, Toronto, London und Sidney 1968
- Bledjian, F. Ergebnisse und neuere theoretische Ansätze über die Wirkungen der Massenkommunikation auf die Einstellungen der Rezipienten, Dissertationsdruck, Nürnberg 1969
- Bogardus, E. S. Measuring Social Distance, J. Appl. Sociol. 9, 1925, abgedruckt in: M. Fishbein (Hrsg.): Readings in Attitude Theory and Measurement, New York, London und Sydney 1967
- Boudon, R. Le modèle des classes latentes, Revue Française de Sociologie 3, 1962
- Brown, R. Models of Attitude Change, in: New Directions in Psychology I, New York, Chicago, San Francisco, Toronto und London 1962
- Burros, R. H. The Application of the Method of Paired Comparisons to the Study of Reaction Potential, Psychol. Rev. 58, 1951
- Burros, R. H. The Estimation of the Discriminal Dispersion in the Method of Successive Intervals, Psychometrika 20, 1955

- Campbell, E. Q. Scale and Intensity Analysis in the Study of Attitude Change, Public Opinion Quarterly 26, 1962
- Campbell, N. R. What is Science, New York 1921
- Campbell, N. R. An Account of the Principles of Measurements and Calculations, London 1928
- Carnap, R. Elementary and Abstract Terms, in: A. Danto und S. Morgenbesser (Hrsg.): Philosophy of Science, New York 1960, (Abdruck der Abschnitte 24 und 25 aus R. Carnap: Foundations of Logic and Mathematics, Encyclopedia of Unified Science, Bd. 3, Chicago 1937)
- Carnap, R. Beobachtungssprache und theoretische Sprache, in: Dialektika 12, 1958
- Clark, K. E.
Kriedt, P. H. An Application of Guttman's New Scaling Techniques to an Attitude Questionnaire, Educ. Psychol. Measmt. 8, 1948
- Cohen, A. R. Attitude Change and Social Influence, New York und London 1964
- Coombs, C. H. Psychological Scaling without a Unit of Measurement, Psychol. Rev. 57, 1950
- Coombs, C. H. A Theory of Psychological Scaling, Engng. Res. Inst. Bull. Nr. 34, Ann Arbor 1952
- Coombs, C. H. Theory and Methods of Social Measurement, in: L. Festinger und D. Katz (Hrsg.): Research Methods in the Behavioral Sciences, New York, Chicago, San Francisco, Toronto und London 1953
- Coombs, C. H. A Theory of Data, New York, London und Sydney 1964
- Coombs, C. H.
Kao, R. C. On a Connection between Factor Analysis and Multidimensional Unfolding, in: H. Gulliksen und S. Messick (Hrsg.): Psychological Scaling: Theory and Applications, New York und London 1960

- Couch, A.
Keniston K. Yeasayers and Naysayers: Agreeing Response Set as a Personality Variable, J. Abn. Soc. Psychol. 60, 1960
- Cronbach, L. J. Essentials of Psychological Testing, New York, Evanston, London und Tokyo, 2. Aufl. 1961, 1. Aufl. 1949
- Cronbach, L. J.
Meehl, P. E. Construct Validity in Psychological Tests, Psychol. Bull. 52, 1955
- Davis, F. B. Item Selection Techniques, in: E. F. Lindquist (Hrsg.): Educational Measurement, Washington 1951
- Davis, F. B. Item Analysis in Relation to Educational and Psychological Testing, Psychol. Bull. 49, 1952
- Diederich, G. W.
Messick, S. J.
Tucher, L. R. A General Least Squares Solution for Successive Intervals, Psychometrika 22, 1957
- Edwards, A. L. A Critique of "Neutral Items" in Attitude Scales Constructed by the Method of Equal-Appearing Intervals, Psychol. Rev. 53, 1946
- Edwards, A. L. On Guttman's Scale Analysis, Educ. Psychol. Measmt. 8, 1948
- Edwards, A. L. The Scaling of Stimuli by the Method of Successive Intervals, J. Appl. Psychol. 36, 1952
- Edwards, A. L. A Technique for Increasing the Reproducibility of Cumulative Attitude Scales, J. Appl. Psychol. 40, 1956
- Edwards, A. L. Techniques of Attitude Scale Construction, New York 1957
- Edwards, A. L.
Kenney, K. C. A Comparison of the Thurstone and Likert Techniques of Attitude Scale Construction, J. Appl. Psychol. 30, 1946, abgedruckt in: M. Fishbein (Hrsg.): Readings in Attitude Theory and Measurement, New York, London und Sydney 1967

- Edwards, A. L.
Kilpatrick, F. P. A Technique for the Construction of
Attitude Scales, J. Appl. Psychol.
32, 1948
- Edwards, A. L.
Kilpatrick, F. P. Scale Analysis and the Measurement
of Social Attitudes, Psychometrika 13,
1948
- Edwards, A. L.
Thurstone, L. L. An Internal Consistency Check for
Scale Values Determined by the
Method of Successive Intervals,
Psychometrika 17, 1952
- Eysenck, H. J.
Crown, S. An Experimental Study in Opinion-
Attitude Methodology, Int. J. Opin.
and Attitude Res. 3, 1949
- Fehrer, E. Shifts in Scale Values of Attitude
Statements as a Function of the
Composition of the Scale, J. Exp.
Psychol. 42, 1955
- Ferguson, L. W. The Influence of Individual Attitudes
on Construction of an Attitude Scale,
J. Soc. Psychol. 6, 1935
- Ferguson, L. W. The Requirements of an Adequate
Attitude Scale, Psychol. Bull. 36,
1939
- Ferguson, L. W. A Study of the Likert Technique of
Attitude Scale Construction,
J. Soc. Psychol. 13, 1941
- Festinger, L. The Treatment of Qualitative Data by
"Scale Analysis", Psychol. Bull. 44,
1947
- Ford, R. N. A Rapid Scoring Procedure for Scaling
Attitude Questions, Public Opinion
Quarterly 14, 1950
- Garner, W. R.
Hake, H. W. The Amount of Information in Absolute
Judgments, Psychol. Rev. 58, 1951
- Gerber, C. P.
Stosberg, M. Die Massenmedien und die Organisation
politischer Interessen, befindet
sich im Druck

- Ghiselli, E. E. Theory of Psychological Measurement, New York, San Francisco, Toronto und London 1964
- Gibson, W. A. An Extension of Anderson's Solution for the Latent Structure Equations, Psychometrika 20, 1955
- Goodenough, W. H. A Technique for Scale Analysis, Educ. Psychol. Measmt. 4, 1944
- Goude, G. On Fundamental Measurement in Psychology, Stockholm, Göteborg und Uppsala 1962
- Granneberg, R. T. The Influence of Individual Attitude and Attitude-Intelligence upon Scale Values of Attitude Items, Amer. Psychol. 10, 1955
- Green, B. F. A General Solution for the Latent Class Model of Latent Structure Analysis, Psychometrika 16, 1951
- Green, B. F. Attitude Measurement, in: G. Lindzey (Hrsg.): Handbook of Social Psychology, 1. Bd., Reading, Mass. und London 1954
- Guilford, J. P. The Method of Paired Comparisons as a Psychometric Method, Psychol. Rev. 35, 1928
- Guilford, J. P. The Computation of Psychological Values from Judgments in Absolute Categories, J. Exp. Psychol. 22, 1938
- Guilford, J. P. The Phi Coefficient and Chi Square as Indices of Item Validity, Psychometrika 6, 1941
- Guilford, J. P. Psychometric Methods, 2. Aufl., New York, Toronto und London 1954
- Gulliksen, H. Theory of Mental Tests, New York, London und Sidney 1950
- Gulliksen, H. A Least Squares Solution for Successive Intervals Assuming Unequal Standard Deviations, Psychometrika 19, 1954

- Guttman, L. The Quantification of a Class of Attributes: A Theory and Method for Scale Construction, in: P. Horst u.a.: The Prediction of Personal Adjustment, New York 1941
- Guttman, L. A Basis for Scaling Qualitative Data, Amer. Soc. Rev. 9, 1944
- Guttman, L. Questions and Answers about Scale Analysis, Research Branch, Information and Education Division, Army Service Forces, Report D-2, 1945
- Guttman, L. On Festinger's Evaluation of Scale Analysis, Psychol. Bull. 44, 1947
- Guttman, L. The Cornell Technique for Scale and Intensity Analysis, Educ. Psychol. Measmt. 7, 1947
- Guttman, L. The Basis for Scalogram Analysis, in: S. A. Stouffer u.a.: Measurement and Prediction, Princeton, N. J., 1950
- Guttman, L. Problems of Reliability, in: S. A. Stouffer u.a.: Measurement and Prediction, Princeton, N. J., 1950
- Guttman, L.
 Suchman, E. A. Intensity and a Zero Point for Attitude Analysis, Amer. Soc. Rev. 12, 1947
- Harding, J.
 Hoguefe, R. Attitudes of White Department Store Employees toward Negro Co-Workers, J. Social Issues 8, 1952
- Hartley, E. L.
 Hartley, R. E. Die Grundlagen der Sozialpsychologie, Berlin 1955
- Hays, D. G.
 Borgatta, E. F. An Empirical Comparison of Restricted and General Latent Distance Analysis, Psychometrika 19, 1954
- Hays, W. L.
 Bennett, J. F. Multidimensional Unfolding: Determining Configuration from Complete Rank Order Preference Data, Psychometrika 26, 1961
- Hempel, C. G. Fundamentals of Concept Formation in Empirical Science, in: O. NEURATH, u.a. (Hrsg.): International Encyclopedia of Unified Science, Bd. 2, Nr. 7, Chicago, London und Toronto 1952

- Hevner, Kate An Empirical Study of Three Psycho-
physical Methods, J. Gen. Psychol. 4,
1930
- Hicks, J. M.
Campbell, D. T. Zero-point Scaling as Affected by
Social Object, Scaling Method, and
Context, J. Pers. Soc. Psychol. 2,
1965
- Hill, R. J. A Note on Inconsistency in Paired
Comparison Judgments, Am. Soc. Rev.
18, 1953
- Hinckley, E. D. The Influence of Individual Opinion
on Construction of an Attitude Scale,
J. Soc. Psychol. 3, 1932
- Horst, P. A Method of Determining the Absolute
Affective Value of a Series of
Stimulus Situations, J. Educ.
Psychol. 23, 1932
- Hovland, C. I.
Sherif, M. Judgmental Phenomena and Scales of
Attitude Measurement: Item
Displacement in Thurstone Scales,
J. Abnorm. Soc. Psychol. 47, 1952,
abgedruckt in: M. Fishbein (Hrsg.):
Readings in Attitude Theory and
Measurement, New York, London und
Sydney 1967
- Hyman, H. Survey Design and Analysis,
New York 1955
- Jurgensen, C. E. Table for Determining Phi Coefficients,
Psychometrika 12, 1947
- Kahn, L. H.
Bodine, A. J. Guttman Scale Analysis by Means of
IBM Equipment, Educ. Psychol. Measmt.
11, 1951
- Kaplan, A. The Conduct of Inquiry, San Francisco
1964
- Kelley, H. H.
Hovland, C. I.
Schwartz, M.
Abelson, R. P. The Influence of Judges' Attitudes
in Three Methods of Attitude Scaling,
J. Soc. Psychol. 42, 1955
- Kerlinger, F. N. Foundations of Behavioral Research,
New York, Chicago, San Francisco,
Toronto und London 1964

- Klapper, J. T. The Effects of Mass Communication, Glencoe, Ill., 1960
- König, R. (Hrsg.) Das Interview, Praktische Sozialforschung I, 4. Aufl., Köln und Berlin 1965
- Krech, D.
Crutchfield, R. S.
Ballachey, E. L. Individual in Society, New York, San Francisco, Toronto, London und Tokyo 1962
- Kriedt, P. H.
Clark, K. E. "Item Analysis" versus "Scale Analysis", J. Appl. Psychol. 33, 1949
- Kuhn, T. S. Die Struktur wissenschaftlicher Revolutionen, Frankfurt 1967 (zuerst unter dem Titel: The Structure of Scientific Revolutions, Chicago 1962)
- Lazarsfeld, P. F. The Interpretation and Computation of Some Latent Structures, in: S. A. Stouffer u.a.: Measurement and Prediction, Princeton, N. J., 1950
- Lazarsfeld, P. F. The Logical and Mathematical Foundation of Latent Structure Analysis, in: S. A. Stouffer u.a.: Measurement and Prediction, Princeton, N. J., 1950
- Lazarsfeld, P. F. A Conceptual Introduction to Latent Structure Analysis, in: P. F. Lazarsfeld (Hrsg.): Mathematical Thinking in the Social Sciences, Glencoe, Ill., 1954
- Lazarsfeld, P. F. Latent Structure Analysis, in: S. Koch (Hrsg.): Psychology, A Study of a Science, Bd. 3, New York, Toronto und London 1959
- Lazarsfeld, P. F. Notes on the History of Quantification in Sociology - Trends, Sources and Problems, in: H. Woolf (Hrsg.): Quantification - A History of the Meaning of Measurement in the Natural and Social Sciences, Indianapolis und New York 1961
- Lazarsfeld, P. F.
Henry, N. W. Latent Structure Analysis, Boston, New York, Atlanta, Geneva, Ill., Dallas und Palo Alto 1968

- Likert, R. A Technique for the Measurement of Attitudes, Arch. Psychol. Nr. 140, 1932
- Likert, R.
Roslow, S.
Murphy, G. A Simple and Reliable Method of Scoring the Thurstone Attitude Scales, J. Soc. Psychol. 5, 1934
- Loevinger, J. The Technic of Homogeneous Tests Compared with Some Aspects of "Scale Analysis" and Factor Analysis, Psychol. Bull. 45, 1948
- Lord, F. M. The Relation of Test Score to the Trait Underlying the Test, Educ. Psychol. Measmt. 13, 1953, abgedruckt in: P. F. Lazarsfeld und N. W. Henry (Hrsg.): Readings in Mathematical Social Science, Chicago 1966
- Lorge, I. The Thurstone Attitude Scales: I. Reliability and Consistency of Rejection and Acceptance, J. Soc. Psychol. 10, 1939
- Malewski, A. Verhalten und Interaktion, Tübingen 1967
- Manis, M. The Interpretation of Opinion Statements as a function of Recipient Attitude, J. Abn. Soc. Psychol. 60, 1960
- Manis, M. The Interpretation of Opinion Statements as a Function of Message Ambiguity and Recipient Attitude, J. Abn. Soc. Psychol. 63, 1961
- Manis, M. The Interpretation of Opinion Statements as a Function of Recipient Attitude and Source Prestige, J. Abn. Soc. Psychol. 63, 1961
- Manis, M. Comment on Upshaw's "Own Attitude as an Anchor in Equal-Appearing Intervals", J. Abn. Soc. Psychol. 68, 1964
- Manis, M. Rejoinder to Upshaw's Reply, J. Abn. Soc. Psychol. 68, 1964

- Margenau, H. The Nature of Physical Reality,
New York, Toronto und London 1950
- McElwain, D. W.
Keats, J. A. Multidimensional Unfolding: Some
Geometrical Solutions, Psychometrika
26, 1961
- Merton, R. K. Social Theory and Social Structure,
Glencoe 1957
- Miller, D. C. Handbook of Research Design and Social
Measurement, New York 1964
- Mosier, C. J. A Modification of the Method of
Successive Intervals, Psychometrika 5,
1940
- Mosteller, F. A Theory of Scalogram Analysis Using
Noncumulative Types of Items,
Report Nr. 9, Laboratory of Social
Relations, Havard University 1949
- Mosteller, F. Remarks on the Method of Paired
Comparisons: I. The Least Squares
Solution Assuming Equal Standard
Deviations and Equal Correlations,
Psychometrika 16, 1951
- Mosteller, F. Remarks on the Method of Paired
Comparisons: II. The Effect of an
Aberrant Standard Deviation when
Equal Standard Deviations and Equal
Correlations Are Assumed,
Psychometrika 16, 1951
- Mosteller, F. Remarks on the Method of Paired
Comparisons: III. A Test of
Significance for Paired Comparisons
when Equal Correlations and Equal
Standard Deviations Are Assumed,
Psychometrika 16, 1951
- Murphy, G.
Likert, R. Public Opinion and the Individual,
New York 1937
- Nystrom, G. H. The Measurement of Filipino Attitudes
toward America by the Use of the
Thurstone Technique, J. Soc. Psychol.
4, 1933

- Osgood, C. E.
Succi, G. J.
Tannenbaum, P. H. The Measurement of Meaning,
Urbana, Ill., 1957
- Payne, S. L. The Art of Asking Questions,
Princeton, N. J., 1951
- Phillips, B. S. Social Research, New York und
London 1966
- Pinter, R.
Forlano, G. The Influence of Attitude upon Scaling
of Attitude Items, J. Soc. Psychol. 8,
1937
- Popper, K. R. Die Zielsetzung der Erfahrungswissen-
schaft, in: H. Albert (Hrsg.):
Theorie und Realität, Tübingen 1964
- Popper, K. R. Das Elend des Historizismus,
Tübingen 1965
- Popper, K. R. Logik der Forschung, 2. deutsche
Aufl., Tübingen 1966
- Prothro, E. T. The Effect of Strong Negative
Attitudes on the Placement of Items
in a Thurstone Scale, J. Soc.
Psychol. 41, 1955
- Prothro, E. T. Personal Involvement and Item
Displacement on Thurstone Scales,
J. Soc. Psychol. 45, 1957
- Remmers, H. H. Introduction to Opinion and Attitude
Measurement, New York 1954
- Riley, White M.
Riley, J. W. Jr.
Toby, J. Sociological Studies in Scale
Analysis, New Brunswick, N. J., 1954
- Rimoldi, H. J. A.
Hormaeche, M. The Law of Comparative Judgment in
the Successive Intervals and Graphic
Rating Scale Methods, Psychometrika
20, 1955
- Rogers, J. L. Prospective Teacher's Attitudes
Toward Freedom of Information,
Journalism Quarterly 32, 1955, z.T.
abgedruckt in: M. E. SHAW und J. M.
WRIGHT: Scales for the Measurement of
Attitudes, New York, St. Louis,
San Francisco, Toronto, London und
Sydney 1967, S. 541 ff.

- Rosander, A. C. The Spearman-Brown Formula in
Attitude Scale Construction,
J. Exp. Psychol. 19, 1936
- Roth, E. Einstellung als Determination
individuellen Verhaltens,
Göttingen 1967
- Rühl, M. Die Zeitungsredaktion als organi-
siertes soziales System, Bielefeld 1969
- Saffir, M. A. A Comparative Study of Scales
Constructed by Three Psychophysical
Methods, Psychometrika 2, 1937
- Sagi, P. C. A Statistical Test for the
Significance of the Coefficient of
Reproducibility, Psychometrika 24,
1959
- Scheuch, E. K. Skalierungsverfahren in der Sozial-
forschung, in: R. König (Hrsg.):
Handbuch der empirischen Sozialfor-
schung, Bd. 1, 2. Aufl.,
Stuttgart 1967
- Scheuch, E. K. Skalierung, in: R. König (Hrsg.):
Handbuch der empirischen Sozial-
forschung, Bd. 1, 2. Aufl.,
Stuttgart 1967, im Anhang
- Scheuch, E. K. Skalierungsverfahren als Instrument
der Sozialforschung, Meisenheim a.G.
1967, bei Abschluß der Arbeit noch
nicht erschienen, zitiert bei: E. K.
Scheuch: Skalierung, in: R. König
(Hrsg.): Handbuch der empirischen
Sozialforschung, Bd. 1, 2. Aufl.,
Stuttgart 1967, im Anhang
- Schuessler, K. F. A Note on Statistical Significance
of Scalogram, Sociometry 24, 1961
- Scott, W. A. Attitude Measurement, in: G. Lindzey
und E. Aronson (Hrsg.): The Handbook
of Social Psychology, Bd. 2, 2. Aufl.,
Reding, Mass., Menlo Park, Calif.,
London, Don Mills, Ontario, 1968
- Seashore, R. H.
Hevner, K. A Time-saving Device for the
Construction of Attitude Scales,
J. Soc. Psychol. 4, 1933

- Selltiz, C.
Jahoda, M.
Deutsch, M.
Cook, S. W. Research Methods in Social Relations,
New York, Chicago, San Francisco und
Toronto 1959
- Shaw, M. E.
Wright, J. M. Scales for the Measurement of
Attitudes, New York, St. Louis,
San Francisco, Toronto, London und
Sydney 1967
- Shepard, R. N. Extracting Latent Structure from
Behavioral Data, Proceedings of the
1964s Symposium on Digital Computing,
January 30-31, 1964, at Bell Telephone
Laboratories, Holmdel
- Shuford, E. H.
Jones, L. V.
Bock, R. D. A Rational Origin Obtained by the
Method of Contingent Paired
Comparisons, Psychometrika 25, 1960
- Siegel, S. A Method for Obtaining an Ordered
Metric Scale, Psychometrika 21, 1956
- Sixtl, F. Meßmethoden der Psychologie,
Weinheim 1967
- Solomon, H. (Hrsg.) Studies in Item Analysis and
Prediction, Stanford 1961
- Stephenson, W. The Study of Behavior, Q-Technique
and Its Methodology, Chicago und
London 1953
- Stevens, S. S. Mathematics, Measurement, and
Psychophysics, in: S. S. Stevens
(Hrsg.): Handbook of Experimental
Psychology, New York, London und
Sydney 1951
- Stosberg, K. Die Selbsteinschätzung der Rezipienten
und die Kommunikationsdiskrepanz als
Determinanten des Einstellungswandels
im Massenkommunikationsprozeß, in
Vorbereitung
- Stouffer, S. A. An Overview of the Contributions to
Scaling and Scale Theory, in: S. A.
Stouffer u.a.: Measurement and Predic-
tion, Princeton, N. J., 1950
- Stouffer, S. A. u.a. Measurement and Prediction, Princeton,
N. J., 1950, Paperback Ausgabe 1966

- Stouffer, S. A. Scaling Concepts and Scaling Theory, in: M. Jahoda, M. Deutsch und S. W. Cook (Hrsg.): Research Methods in Social Relations, Bd. 2, New York 1951
- Stouffer, S. A.
 Borgatta, E. F.
 Mays, D. G.
 Henry, A. F. A Technique for Improving Cumulative Scales, Public Opinion Quarterly 16, 1952
- Suchman, E. A. The Intensity Component in Attitude and Opinion Research, in: S. A. Stouffer u.a.: Measurement and Prediction, Princeton, N. J., 1950
- Suchman, E. A. The Scalogram Board Technique, in: S. A. Stouffer u.a.: Measurement and Prediction, Princeton, N. J., 1950
- Suchman, E. A. The Utility of Scalogram-Analysis, in: S. A. Stouffer u.a.: Measurement and Prediction, Princeton, N. J., 1950
- Suppes, P.
 Zinnes, J. L. Basic Measurement Theory, in: R. D. Luce, R. R. Bush und E. Galanter (Hrsg.): Handbook of Mathematical Psychology, New York und London 1963
- Tannenbaum, P. H. Attitudes Toward Source and Concept as Factors in Attitude Change through Communications, Unveröffentlichte Dissertation, Illinois 1953
- Thurstone, L. L. Psychophysical Analysis, Am. Psychol. 38, 1927
- Thurstone, L. L. A Law of Comparative Judgment, Psychol. Rev. 34, 1927
- Thurstone, L. L. A Scale for Measuring Attitude toward the Movies, J. Educ. Res. 22, 1930, z.T. abgedruckt in: M. E. Shaw und J. M. Wright: Scales for the Measurement of Attitudes, New York, St. Louis, San Francisco, Toronto, London und Sidney 1967
- Thurstone, L. L. Stimulus Dispersion in the Method of Constant Stimuli, J. Exp. Psychol. 15, 1932

- Thurstone, L. L. The Measurement of Values,
Chicago 1959
- Thurstone, L. L.
Chave, E. J. The Measurement of Attitude,
Chicago 1929
- Thurstone, L. L.
Jones L. V. The Rational Origin for Measuring
Subjective Values, in: L. L.
Thurstone: The Measurement of
Values, Chicago 1959
- Torgerson, W. S. A Law of Categorical Judgment,
Amer. Psychol. 9, 1964
- Torgerson, W. S. Theory and Methods of Scaling,
New York, London und Sydney 1958
- Tucker, L. R. Description of Paired Comparison
Preference Judgments by a
Multidimensional Vector Model,
Educational Testing Service,
Princeton, N. J., 1955
- Uhrbrock, R. S. Attitude of 4430 Employees,
J. Soc. Psychol. 5, 1934
- Upshaw, H. S. Own Attitude as an Anchor in
Equal-Appearing Intervals,
J. Abn. Soc. Psychol. 64, 1962
- Upshaw, H. S. A Linear Alternative to Assimilation-
Contrast: A Reply to Manis, J. Abn.
Soc. Psychol. 68, 1964
- Upshaw, H. S. The Effect of Variable Perspectives
on Judgment of Opinion Statements
for Thurstone Scales: Equal-Appearing
Intervals, J. Pers. Soc. Psychol. 2,
1965
- Upshaw, H. S. Attitude Measurement, in: H. M.
Blalock Jr. und A. B. Blalock (Hrsg.):
Methodology in Social Research, New
York, St. Louis, San Francisco,
Toronto, London und Sydney 1968

- Ward, C. D. Ego Involvement and the Absolute Judgment of Attitude Statements, J. Pers. Soc. Psychol. 2, 1965
- Ward, C. D. Attitude and Involvement in the Absolute Judgment of Attitude Statements, J. Pers. Soc. Psychol. 4, 1966
- Webb, S. C. A Generalized Scale for Measuring Interest in Science Subjects, Educ. Psychol. Measmt. 11, 1951
- White, B. W.
Saltz, E. Measurement of Reproducibility, Psychol. Bull. 54, 1957
- Wilke, W. H. An Experimental Comparison of the Speech, the Radio, and the Printed Page as Propaganda Devices, Arch. Psychol. 169, 1934, z.T. abgedruckt in: M. E. Shaw und J. M. Wright: Scales for the Measurement of Attitudes, New York, St. Louis, San Francisco, Toronto, London und Sydney 1967
- Wohlwill, J. The Measurement of Scalability for Non-Cumulative Items, Educ. Psychol. Measmt. 23, 1963
- Woolf, H. (Hrsg.) Quantification. A History of the Meaning of Measurement in the Natural and Social Sciences, Indianapolis und New York 1961

L e b e n s l a u f

Am 20. Juli 1940 wurde ich als einziges Kind des städtischen Verwaltungsrates a.D. Oskar Stosberg und seiner Ehefrau Else, geb. Böhl, in Hückeswagen (Rhein-Wupper-Kreis) geboren. Dort besuchte ich von 1947 an die Volksschule und wechselte 1951 zur Realschule über. 1953 trat ich in das neusprachliche Röntgen-Gymnasium in Remscheid-Lennep ein; an dieser Schule bestand ich im Frühjahr 1960 das Abitur.

Im Sommersemester 1960 immatrikulierte ich mich an der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität in Köln. Dort legte ich im Juni 1965 die Diplomprüfung für Volkswirte ab. Im Anschluß daran bekam ich eine Stelle als Assistent am Seminar für Soziologie der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität Erlangen-Nürnberg. Seit Beginn des Jahres 1968 bin ich Mitglied des Sozialwissenschaftlichen Forschungszentrums der Universität.

Im September 1965 heiratete ich die Diplom-Volkswirtin Krista Beitzel aus Remscheid.

Offsetdruck-Fotodruck

J. Hogl, 8520 Erlangen Hauptstr. 109, (Nähe Martin-Luther-Platz) Tel. 22990
